SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

SERIES A

VOLUME 25, PART 1

EDITORS

P. C. MAHALANOBIS

C. R. RAO

JULY, 1963

STATISTICAL PUBLISHING SOCIETY
CALCUTTA

SANKHYĀ, SERIES A, VOL. 25, PART 1, 1963

CONTENTS

SANKHYĀ : THE INDIAN JOURNAL OF STATISTICS

CONTENTS OF SERIES A, VOLUME TWENTYFIVE, 1963

Papers

IK 5

Addenda

Corrigenda

Author Index

BIOMETRIKA

Volume 50, Parts 1 and 2 **June, 1963**

CONTENTS

The subscription, *payable in advance*, is 54/- (or \$8.00), per volume (including postage). Cheques should be made payable to *Biometrika*, crossed "*A/c Biometrika Trust*" and sent to the Secretary, Biometrika Office, Department of Statistics, University College, London, W. C. 1. All foreign cheques must be drawn on a Bank having a London agency.

Issued by THE BIOMETRIKA OFFICE, University College, London.

JOURNAL OF THE

ROYAL STATISTICAL SOCIETY

Series B (Methodological)

40*s* each Part. Annual Subscription £3.2*s* (US \$9). Post free

CONTENTS OF VOL. 25, PART 1, 1963

Obtainable through any bookseller or direct from the

ROYAL STATISTICAL SOCIETY

21 Bentinck Street, London, W. 1.

THE ANNALS OF MATHEMATICAL STATISTICS

(Founded by H. C. Carver)

The Official Journal of the Institute of Mathematical Statistics

CONTENTS OF VOLUME 34, NO. 2, JUNE, 1963

Subscription Rate: $ 10.00 per year

Address orders for subscriptions and back numbers to the Secretary, Institute of Mathematical Statistics, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

ECONOMETRICA

Vol. 31, No. 1-2 **January-April, 1963**

CONTENTS

Subscription price for nonmembers : $ 12.00 per year in the United States and Canada; $ 10.00 per year in all other Countries. Single copies $ 3.00

Graduate Students recommended by a faculty member : $ 4.00 per year.

Subscriptions should be sent to

The Secretary

ECONOMETRIC SOCIETY

Box 1264, Yale Station, New Haven, Connecticut, U.S.A.

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Editors: P. C. MAHALANOBIS, C. R. RAO

SERIES A, VOL. 25 JULY 1963 PART 1

INTRODUCING VOLUME TWENTYFIVE

1. The first number of *Sankhyā* : The Indian Journal of Statistics was published exactly thirty years, or one generation ago, in June 1933. In starting the twentyfifth volume I have much pleasure in announcing that Dr. C. Radhakrishna Rao would be associated with me as editor in future. The pleasure is all the greater because he is my former pupil; he attended my lectures as a student in the post-graduate course in statistics which was started for the first time in India in the University of Calcutta in 1941. After taking his M.A. degree in statistics, he joined the Indian Statistical Institute in 1943, and since then has been my colleague for twenty years. He has been, in actual fact, in charge of most of the editorial work for many years. I look forward to the day when he would assume full responsibility.

2. Having attained on this day the age of seventy, I may take the opportunity of looking both back and forward, a little. My mind goes back to 1933 when we were busy sending articles to the press for the first number of *Sankhyā*. I had started working on the multivariate distance several years earlier, and had examined a portion of a large volume of individual measurements of various anthropometric characters for a large number of castes and tribes in North India which had been published in 1891 by H. H. Risley. The measurements had been taken by the same small group of observers; and were, therefore, suitable for purposes of comparisons between castes and tribes. Karl Pearson had, however, condemned the material much earlier because of the discrepancies he had found in the average values given by Risley. After detailed scrutiny, I reached the conclusion that most of the discrepancies in individual measurements and indices could be traced to easily recognisable copying or printing mistakes, use of wrong figures taken from adjoining rows or columns, mistakes in entering index tables, or obvious arithmetical slips like a displacement of a decimal point in calculations. Out of 142 discrepancies in individual values, 133 could be corrected and corroborated with practical certainty by cross-checks with appropriate index numbers; in eight cases, the corrections were plausible although they could not be confirmed, while only one single measurement was really doubtful and had to be rejected out of a total of 12,197 individual measurements and a total of 8,600 indices given by Risley. The real defect in Risley's data had occurred during the calculation of average values; the primary data of individual measurements given by him could be used with safety, especially after applying the corrections which I had used.

3. I thought it would be useful to publish the revised values of Risley's data together with the detailed evidence in support of the corrections made by me.

My young colleagues, who were only three or four at that time, were strongly opposed to the publication of this paper. The scrutiny and reconciliation of discrepancies which I had carried out, they felt, could not be considered to be scientific work at all; they were eager to prevent the 'Professor' from exposing himself to ridicule in advanced countries by publishing this paper. However, due to a streak of contrariness and obstinacy, I printed this article in the first issue of the journal.[1]

4. It can be easily imagined with what joy and encouragement I read the following lines in a letter dated 14 August 1933 from Ronald Aylmer Fisher :—

> "You are most heartily to be congratulated on the new Journal, and very especially on your own contributions. The work on Risley's data will be most valuable. I shall hope to hear and read more of your contributions as time goes on. It is a splendid start."

It was characteristic of the most eminent statistician of the present age to have selected the paper on Risley's data for special mention.

5. Fisher himself had said somewhere that the first responsibility of a statistician is to cross-examine his data. I remember the vivid description he gave me, during his first visit to the Indian Statistical Institute in 1937, about his investigations in the monastry in Austria where Gregor Mendel had carried out his experiments on the inheritance of characters in sweet peas. Mendel had announced in his last scientific publication that he would publish a in another paper his results on three factor segregation, but did not do so. Fisher had an almost irresistible urge to find out why Mendel ceased publication. Searching through old records, Fisher traced the original observations which Mendel had intended to use for his unpublished paper, and found that there was perfect agreement between observed and expected results. Fisher surmised that such agreement had raised a suspicion in Mendel's mind that his assistant, who had been helping him in these experiments, had deliberately changed the records to make them agree with expectations; Mendel had refrained from publishing the results as he could not guarantee their accuracy.[2]

6. My mind also goes back to the day when I had the good fortune to establish contact with R. A. Fisher. In 1923 I was working as Meteorologist in Calcutta in addition to teaching physics in the Calcutta University. Fisher was engaged in his researches on the design of experiments at Rothamsted Experimental Station, Harpenden. I had no connexion with agricultural research. By sheer chance, my attention was drawn to the question of "errors" in some agricultural field experiments, (in the form of a series of parallel plots sown with different varieties of rice, repeated in the same order in several blocks). I tried to eliminate, by crude graduation, differences in soil fertility and published a paper in an agricultural journal.[3] Fisher saw this paper and immediately sent me reprints of his early papers on the design of experiments and also the paper on the distribution of the ratio of two variances.

7. While struggling with the analysis of variety trials on paddy, I had begun to appreciate the need of radical improvements in agricultural field experiments. When I read Fisher's papers on this subject. I realized that he had not only solved the problem at a theoretical level but had also supplied the basic tables

[1] A Revision of Risley's Anthropometric Data Relating to the Tribes and Castes of Bengal. *Sankhyā*, **1**, 1933, 76—105.

[2] R. A. Fisher : "Has Mendel's work been rediscovered ?" *Annals of Science* **1**, 115—137, 1936.

[3] Probable error of field experiments in agriculture. *Agri. Jour. Ind.* **20**, 1925, 96.

(for the z-distribution) to facilitate the use of his methods almost in a routine manner. I could also appreciate how great was his achievement. I believe, I can claim to be the first convert to the Fisherian view of statistics; I have also tried to extend his ideas to the design of sample surveys. For me, the discovery of Fisher, nearly forty years ago, was an important factor in deepening my interest in statistics which was further strengthened by the impressions of the memorable day I spent with him at Rothamsted Agricultural Station in 1926 when I met him for the first time.

8. I also recall that it was at Fisher's suggestion (as I came to know much later) that the newly established Imperial (now Indian) Council of Agricultural Research offered me in 1928 an annual grant of Rs. 2,500 (about £ 200, a princely sum for us in those days) to have a research assistant to take up some work in statistics. This grant led the way to the future development of the integrated programme of theoretical research, training, and applied projects which has been a characteristic feature of the Indian Statistical Institute.

9. Fisher came to our Institute on eight occasions. He always stayed in our house in the Institute in Calcutta. This gave me and my young colleagues the opportunity to profit by his stimulating discussions and suggestions. The special needs of an underdeveloped country like India, had made it continually necessary for us to increase the scope of application of statistical methods in widely differing subject fields in natural and social sciences, technology and economic planning. In such developments we received powerful support from R. A. Fisher, who quite early had a clear view of statistics as the new technology of the modern age.[4] He also first formulated, in a precise way, the concept of the Indian Statistical Institute as higher technological institution having an analogous function in respect of statistics, although on a much smaller scale, to that of the higher technological institutions like the Zurich Federal School of Technology or the Massachusetts Institute of Technology, which had been established a hundred years ago to provide an integrated programme of research, training and projects in the field of engineering and technology. In all these ways, Ronald Fisher had exercised more influence than any one else in the shaping of the policy and programme of the Indian Statistical Institute of which *Sankhyā* is the official organ.

10. R. A. Fisher had said somewhere that he had learnt his statistics through computation, I presume, in the dual sense that no theoretical formulae are of any value unless these can be used in numerical terms at a concrete level, and also thta statistician have to do the 'dirty work' of computation with their own hands. On the occasion of his first visit to the Institute in December 1937, he requested me to give him a hand calculating machine. For his seven subsequent visits, a desk calculator always used to be placed in his room in advance; he used such a calculator every day during his last stay in Calcutta up to the middle of February 1962.

11. I am recalling the two points made by him, namely, the need of cross-examining the data and the importance of computational work in statistics, because both these points are of crucial significance in the future development of statistics in India. During the last thirty years or the first generation of this journal, there has been a good deal of advance in the field of statistics in India. Indian statisticians have come to enjoy a good reputation abroad. Paradoxically, the Indian statistical system in respect of flow of factual information is recognized, within and outside the country, to be very weak. This is due to the fact that collection, processing and

[4] Reference to R. A. Fisher's address at the first convocation in 1962; also to Fisher's speech in 1951, quoted by me in an annual review.

publication of official statistics are still treated as administrative matters, subject to the principle of a monopolistic jurisdiction of one single administrative agency for each type of information. The acceptability of statistical data is determined by the status of the authority responsible for their collection or publication, usually without any assessment of the reliability of the information.

12. The only way to improve the quality of official statistics in India is by testing their accuracy in accordance with accepted scientific principles, by using checks and cross-checks provided through multiple observations or through independent sources of information. Sample surveys, with multiple and parallel or inter-penetrating network of sub-samples, provide a speedy and economical way of ascertaining the margin of uncertainty in an objective manner. It has been, therefore, the policy of this journal to attach special importance to methods and practical applications of sample surveys. The need of doing this would continue in future.

13. In the second generation, if I may call it so, of the journal, we feel one of its important tasks must be to foster the growth of the spirit of criticism without which the advancement of science is not possible. In the early years, we used to publish book reviews and also selective reviews and abstracts of statistical papers published elsewhere; much to our regret, we have not been able to maintain these two features. We feel it may be particularly helpful to use reviews of articles and books, in a purposeful manner, to promote the growth of a critical appreciation of the quality of research and of advances in the collection, processing and analysis of statistical information in India. In future, it would be our policy to welcome reviews both of papers published in journals and books.

14. From the very beginning, to help in the advancement of statistics in an underdeveloped country like India, we had adopted certain lines of policy which are somewhat unorthodox, for example, reprinting papers published elsewhere to make them more easily available in India, publishing statistical materials and data for purposes of record, or giving attention to primary work like scrutiny of data and computations. It is our intention to continue these features. At the same time, we shall continue to welcome contributions from all over the world, as we have been doing from the first issue of the journal.

15. I should like to take this opportunity of placing on record my sense of gratitude to the late Narendra Nath Mookerjee, who with characteristic generosity had helped in the publication of the journal from the Art Press, in the early years. I remember the services rendered by my two young colleagues, Subhendu Sekhar Bose and Sudhir Kumar Banerjee, who passed away at a very early age. In recent years, we are grateful for the help we have received from Anikendra Mahalanobis, Krishna Birendra Goswami and Dyutish Ranjan Banerjee in the publication of this journal. We offer our sincere thanks to the workers of the Eka Press, who have always helped ungrudgingly in printing the journal.

16. It was an adventure, even foolhardy, to have started a statistical journal in India thirty years ago when our resources in research and in material equipment were meagre. It is because of the friendly cooperation and support which we have been receiving from within and outside India that this journal could make a place of its own in public esteem. We offer our most sincere thanks to our friends and contributors in India and abroad for their kind cooperation, and seek their continuing support in future.

29 June, 1963

P. C. Mahalanobis

LARGE SAMPLE SEQUENTIAL TESTS FOR COMPOSITE HYPOTHESES

By D. R. COX

Birkbeck College, University of London

SUMMARY. A sequential test for hypotheses involving nuisance parameters is developed from maximum likelihood (m.l.) theory. The procedure is a slight modification of one outlined by Bartlett (1946a). Special cases are discussed.

1. Introduction

Except for a short section in a paper by Bartlett (1946a), previous work on sequential tests when there are nuisance parameters has been restricted to very special problems. Most work has concerned situations where there are simple sufficient statistics and where invariance ideas are applicable to the estimation of the nuisance parameters. Barnard (1946, 1952), Wald (1947, p. 80) and Cox (1952) have explored this from rather different points of view. In appropriate cases, a sequential test can be formed by finding at each step the relevant standard fixed-sample size statistic, and computing the ratio of the densities of this under the two base hypotheses H_1 and H_2. This ratio is then used with the usual stopping limits $(1-\beta)/\alpha$ and $\beta/(1-\alpha)$, where α and β are the probabilities of error under H_1 and H_2. The most important example of this procedure is the sequential t test (Rushton, 1950).

Other results include that of Girshick (1946), who gave a procedure for comparing two populations. However, his procedure often leads to an operating characteristic depending upon an undesirable combination of population parameters. For example, the test for comparing two normal variances σ_1^2 and σ_2^2 has an operating characteristic depending on $1/\sigma_1^2-1/\sigma_2^2$, and this would not usually be what is wanted. Another special procedure is for comparing two binomial probabilities (Wald, 1947, p. 106). This will be discussed in Section 4.

In the present paper, a slight modification is given of Bartlett's procedure, which is based on maximum likelihood (m.l.) theory. A number of special cases are then examined.

We consider, for simplicity, tests corresponding to the Wald likelihood ratio test for comparing two simple hypotheses. Extensions to the comparison of three hypotheses and to the important closed schemes of Armitage (1957) can be made without difficulty.

2. TEST BASED ON MAXIMUM LIKELIHOOD ESTIMATES

Let the distribution of the observations be known except for the unknown parameters (θ, ϕ). Let the two base hypotheses be $H_i : \theta = \theta_i (i = 1, 2)$, the quantity ϕ being a nuisance parameter. The dimensionality of θ and ϕ does not matter; we shall write out formulae as if both are one-dimensional.

After n observations, let $L_n(\boldsymbol{x}_n, \theta, \phi)$ be the log likelihood. If ϕ is known, and equal to ϕ_0 say, the test is based on

$$L_n(\boldsymbol{x}_n, \theta_2, \phi_0) - L_n(\boldsymbol{x}_n, \theta_1, \phi_0), \qquad \dots \quad (2.1)$$

using $A = \log \{(1-\beta)/\alpha\}$ and $B = \log \{\beta/(1-\alpha)\}$ as stopping limits. We deal here with large sample theory in which the log likelihood is expanded as far as quadratic terms. In a rigorous treatment we would consider a sequence of schemes depending on a parameter N in such a way that, as $N \to \infty$, the relevant sample sizes are proportional to N, and θ_1, θ_2 and the true value θ differ by amounts of order $1/\sqrt{N}$. That is, in expanding (2.1) about the true value θ, $\theta_1 - \theta$ and $\theta_2 - \theta$ are to be treated as of order $1/\sqrt{n}$. Thus we have that for $i = 1, 2$,

$$L_n(\boldsymbol{x}_n, \theta_i, \phi_0) = L_n(\boldsymbol{x}_n, \theta, \phi_0) + (\theta_i - \theta) \frac{\partial L_n(\boldsymbol{x}_n, \theta, \phi_0)}{\partial \theta} + \tfrac{1}{2}(\theta_i - \theta)^2 \frac{\partial^2 L_n(\boldsymbol{x}_n, \theta, \phi_0)}{\partial \theta^2},$$

where the last two terms are of order 1 in probability. Thus (2.1) becomes

$$(\theta_2 - \theta_1) \frac{\partial L_n(\boldsymbol{x}_n, \theta, \phi_0)}{\partial \theta} + \tfrac{1}{2}(\theta_2 - \theta_1)(\theta_2 + \theta_1 - 2\theta) \frac{\partial^2 L_n(\boldsymbol{x}_n, \theta, \phi_0)}{\partial \theta^2}. \qquad \dots \quad (2.2)$$

Now suppose that ϕ is unknown, and that no prior probability distribution of ϕ is available. Let $(\hat{\theta}, \hat{\phi}_n)$ be the maximum likelihood estimate of (θ, ϕ) based on $\boldsymbol{x}_n$. The method of Bartlett (1946a) involved the use of two maximum likelihood estimates of ϕ, one assuming $\theta = \theta_1$, the other that $\theta = \theta_2$. This is, however, unnecessary to the order of approximation being considered here.

It is plausible to consider instead of (2.1)

$$L_n(\boldsymbol{x}_n, \theta_2, \hat{\phi}) - L_n(\boldsymbol{x}_n, \theta_1, \hat{\phi}). \qquad \dots \quad (2.3)$$

Expanding (2.3) about the true point (θ, ϕ), we get

$$(\theta_2 - \theta_1) \frac{\partial L_n(\boldsymbol{x}_n, \theta, \phi)}{\partial \theta} + \tfrac{1}{2}(\theta_2 - \theta_1)(\theta_2 + \theta_1 - 2\theta) \frac{\partial^2 L_n(\boldsymbol{x}_n, \theta, \phi)}{\partial \theta^2}$$

$$+ (\theta_2 - \theta_1)(\hat{\phi} - \phi) \frac{\partial^2 L_n(\boldsymbol{x}_n, \theta, \phi)}{\partial \theta \partial \phi}. \qquad \dots \quad (2.4)$$

Thus the test based on (2.3) is asymptotically equivalent to that when ϕ is known if and only if

$$\frac{1}{n} \frac{\partial^2 L_n(\boldsymbol{x}_n, \theta, \phi)}{\partial \theta \partial \phi} \to 0 \qquad \dots \quad (2.5)$$

in probability. This is the condition that $\hat{\theta}$ and $\hat{\phi}$ are asymptotically independent. Therefore, when (2.5) is satisfied, we can replace ϕ by $\hat{\phi}$ and apply the usual formulae for simple hypotheses. An asymptotic optimum property will hold corresponding to the exact optimum property when ϕ is known. The practical relevance of this optimum property may, however, be slight unless both θ is likely to be near to θ_1 or θ_2, and also the cost of sampling is proportional to the number of observations.

If (2.5) is not satisfied, the last term in (2.4) is comparable with the others, and to neglect the sampling fluctuations in $\hat{\phi}$ is wrong. To simplify (2.4), we make the usual substitutions of maximum likelihood theory writing

$$-\frac{\partial^2 L_n(\boldsymbol{x}_n, \theta, \phi)}{\partial\theta^2} \sim nI_{\theta\theta}, -\frac{\partial^2 L_n(\boldsymbol{x}_n, \theta, \phi)}{\partial\phi^2} \sim nI_{\phi\phi}, -\frac{\partial^2 L_n(\boldsymbol{x}_n, \theta, \phi)}{\partial\theta\partial\phi} \sim nI_{\theta\phi}, \quad \ldots \quad (2.6)$$

where the I's are known or can be estimated consistently. In particular, if the observations are independently and identically distributed with density function $f(x; \theta, \phi)$, we have the usual formulae, such as

$$I_{\theta\theta} = E\left\{-\frac{\partial^2 \log f(x; \theta, \phi)}{\partial\theta^2}\right\}. \quad \ldots \quad (2.7)$$

The maximum likelihood estimates $\hat{\theta}, \hat{\phi}$ satisfy asymptotically the equations

$$I_{\theta\theta}\,(\hat{\theta}-\theta)+I_{\theta\phi}\,(\hat{\phi}-\phi) = \frac{1}{n}\,\frac{\partial L_n(\boldsymbol{x}_n, \theta, \phi)}{\partial\theta},$$

$$I_{\theta\phi}\,(\hat{\theta}-\theta)+I_{\phi\phi}\,(\hat{\phi}-\phi) = \frac{1}{n}\,\frac{\partial L_n(\boldsymbol{x}_n, \theta, \phi)}{\partial\phi}. \quad \ldots \quad (2.8)$$

We can now express (2.4) in terms of $\hat{\theta}, \hat{\phi}$. In fact (2.4) is asymptotically equivalent to $nI_{\theta\theta}\,(\theta_2-\theta_1)\{\hat{\theta}-\frac{1}{2}(\theta_1+\theta_2)\}$, suggesting that the test should be based on the computation of

$$T_n = n\{\hat{\theta}-\tfrac{1}{2}(\theta_1+\theta_2)\}. \quad \ldots \quad (2.9)$$

Now $\qquad E(T_n) = n\{\theta-\frac{1}{2}(\theta_1+\theta_2)\}$, $\operatorname{var}(T_n) = nI^{\theta\theta}$,

where

$$\frac{1}{I^{\theta\theta}} = I_{\theta\theta} - \frac{I_{\theta\phi}^2}{I_{\phi\phi}}. \quad \ldots \quad (2.10)$$

Further T_n can, from (2.8), be expressed as a linear combination with constant coefficients of

$$\frac{\partial L_n(\boldsymbol{x}_n, \theta, \phi)}{\partial\theta} \text{ and } \frac{\partial L_n(\boldsymbol{x}_n, \theta, \phi)}{\partial\phi}. \quad \ldots \quad (2.11)$$

If, for example, the observations are independent and identically distributed, the quantities (2.11) are sums of independent and identically distributed random variables and hence, asymptotically, so is T_n. The same is true much more generally. That is, the stochastic process $\{T_n\}$ is a random walk in which the mean increment per step is $\theta-\frac{1}{2}(\theta_1+\theta_2)$ and the variance per step is $I^{\theta\theta}$. This has in general to be estimated from $(\hat{\theta}, \hat{\phi})$, but this introduces negligible error to the order being considered. Alternatively $I^{\theta\theta}$ could be estimated by substituting $\theta = \frac{1}{2}(\theta_1+\theta_2)$, $\phi = \hat{\phi}$. We can therefore use the theory for normally distributed observations (Wald, 1947, p. 122).

If we set stopping limits at values of T_n corresponding to constant values of the estimated difference of log likelihoods (2.3), the limits should be at

$$\frac{I^{\theta\theta}}{\theta_2-\theta_1} \log\left(\frac{\beta}{1-\alpha}\right), \quad \frac{I^{\theta\theta}}{\theta_2-\theta_1} \log\left(\frac{1-\beta}{\alpha}\right); \qquad \ldots \quad (2.12)$$

the operating characteristic is, in Wald's notation,

$$L(\theta) = \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h}, \qquad \ldots \quad (2.13)$$

where

$$h = \frac{\theta_2+\theta_1-2\theta}{\theta_2-\theta_1}. \qquad \ldots \quad (2.14)$$

Also, the expected sample size is $(\theta \neq \frac{1}{2}\theta_1+\frac{1}{2}\theta_2)$

$$\frac{I^{\theta\theta}\left[L(\theta)\log\left(\frac{\beta}{1-\alpha}\right)+\{1-L(\theta)\}\log\left(\frac{1-\beta}{\alpha}\right)\right]}{(\theta_2-\theta_1)\{\theta-\frac{1}{2}(\theta_2+\theta_1)\}}. \qquad \ldots \quad (2.15)$$

Note that, as pointed out by Bartlett (1946b), the ratio of mean sample sizes when ϕ is (a) known and (b) unknown is the same as the corresponding ratio of asymptotic variances in ordinary maximum likelihood estimation.

There are many procedures asymptotically equivalent to the one sketched above :

(a) instead of the maximum likelihood estimate $\hat{\theta}$ we can use in (2.9) an asymptotically equivalent estimate;

(b) for the consistent estimation of $I^{\theta\theta}$ a variety of methods will often be available;

(c) if the parameter θ is transformed by a monotone transformation to $g(\theta)$, the procedure can be operated in terms of

$$n[g(\hat{\theta})-\tfrac{1}{2}\{g(\theta_1)+g(\theta_2)\}].$$

We shall not consider here (a) and (b) further. As for (c), it seems reasonable to choose θ so that the likelihood function $L_n(x_n, \theta, \phi)$ is as far as possible quadratic with parameters not sharply dependent on (θ, ϕ), that is to try to make $\hat{\theta}$ as nearly as possible normally distributed with constant variance.

For example, suppose that pairs of observations are taken from normal populations of standard deviations σ_1, σ_2, it being required to make, by the present method, a sequential test concerning σ_1/σ_2. It would be reasonable to take $\theta = \log(\sigma_1/\sigma_2)$. The maximum likelihood estimate $\hat{\theta}$ after n pairs of observations, Fisher's z, is nearly normally distributed with variance independent of the parameters, $I^{\theta\theta} = 1$. The procedure based on (2.9) is nearly equivalent to that based on the exact density of the variance ratio distribution (Cox, 1952). Formulae (2.13) and (2.15) for the operating characteristic and expected sample size are new.

3. Tests for the mean of a normal distribution

Consider the sampling of a single normal population of unknown mean μ and variance σ^2. If we are interested in the mean of the population, there are two main formulations, corresponding to defining the parameters θ, ϕ by

$$\text{(i)} \quad \theta = \mu/\sigma, \ \phi = \sigma;$$

$$\text{(ii)} \quad \theta = \mu, \quad \phi = \sigma.$$

The first is the form adopted in work on the sequential t test, and the hypotheses $\theta = \theta_1, \theta_2$ are, effectively, hypotheses about the probability of obtaining a positive observation. The second formulation would be appropriate if certain absolute changes in the mean are of interest, irrespective of the value of σ. This might be the case in measurements on an extensive property.

The sequential procedure in case (ii) is obtained by using the usual estimate s to replace σ in the procedure with σ known (Wald, 1947, p. 122). Since s and the sample mean are independent, there is asymptotically no change in the properties of the scheme from those when σ is known.

In case (i), the maximum likelihood estimator $\hat{\theta}$ is, in the usual notation $\bar{x}/s$, and it follows from first principles or from the general theory, that asymptotically

$$\operatorname{var}(\hat{\theta}) = \frac{I^{\theta\theta}}{n} = \frac{\sigma^2}{n}\,(1+\tfrac{1}{2}\theta^2). \qquad \dots \quad (3.1)$$

Thus the procedure defined by (2.9) and (2.12) is to use the statistic

$$n\left\{\frac{\bar{x}}{s} - \tfrac{1}{2}(\theta_1+\theta_2)\right\} \qquad \dots \quad (3.2)$$

with stopping limits at

$$\left\{\frac{1+\frac{1}{2}\frac{\bar{x}^2}{s^2}}{\theta_2-\theta_1}\right\}\left\{\log\left(\frac{\beta}{1-\alpha}\right), \log\left(\frac{1-\beta}{\alpha}\right)\right\}. \qquad \dots \quad (3.3)$$

The mean sample size is, from (2.15) and (3.1), $(1+\frac{1}{2}\theta^2)$ times what it would have been with σ known. The test based on (3.2) and (3.3) is asymptotically equivalent to, and numerically close to, the test in the form given by Rushton (1950) and the National Bureau of Standards' tables (1951).

4. Comparison of binomial proportions

One of the most important sequential problems is the comparison of two binomial proportions. Denote the two possible outcomes of each kind by 0 and 1. We suppose here that the observations are paired, but this is not essential for the tests to be developed below. There are four possible results for a pair of observations, (i) [0, 0], (ii) [0,1], (iii) [1, 0], (iv) [1, 1].

Wald (1947, p. 106) obtained a test by rejecting types (i) and (iv) and examining the proportion of type (iii) among the remaining responses. There are two conditions for this to be completely appropriate (Cox, 1958). First, suppose that in the

i-th pair the odds of a 1 are respectively ϕ_i and $\theta\phi_i$ in the two populations, where the ϕ_i's are nuisance parameters, one for each pair. Secondly, suppose that θ, the odds ratio, is the sole parameter of interest. That is, the probability of a 1 in the i-th pair is $\phi_i/(1+\phi_i)$ in the first group and $\theta\phi_i/(1+\theta\phi_i)$ in the second group. Note that even if this probability model holds, θ is not necessarily the parameter of interest. For example, there may be economic reasons for being concerned with the difference between the overall proportions of 1's in the two groups.

If, however, we observe two independent series of Bernoulli trials, Wald's procedure is not completely efficient (Barnard, 1946; Wald, 1947, p. 108). For it is not based on the sufficient statistics for the problem. From now on, suppose that all trials are independent and that the probabilities of a 1 are λ_1 and λ_2 in the two populations. We can formulate various hypotheses to which the theory of Section 2 can be applied and shall here consider two, corresponding to situations in which (i) the difference of probabilities and (ii) the ratio of odds are the parameters of interest. For (i) the natural parametrization is to take

$$\lambda_1 = \phi, \qquad \lambda_2 = \phi+\theta. \qquad \dots \quad (4.1)$$

For (ii) it is perhaps better to take θ as the difference of log odds rather than the ratio of odds (see the final remarks of Section 2). That is, we write

$$\lambda_1 = \frac{\phi}{1+\phi}, \qquad \lambda_2 = \frac{\phi e^{\theta}}{1+\phi e^{\theta}} \qquad \dots \quad (4.2)$$

In both (4.1) and (4.2) the base hypotheses are $\theta = \theta_1, \theta_2$. Often $\theta_1 = 0$, corresponding to a null hypothesis $\lambda_1 = \lambda_2$.

We can now apply the results of Section 2 in a routine way. After n pairs of observations, let there be $r_n^{(1)}$, $r_n^{(2)}$ 1's in the two groups. The test statistics (2.9) are

case (i) $$r_n^{(2)}-r_n^{(1)}-\tfrac{1}{2}n(\theta_1+\theta_2); \qquad \dots \quad (4.3)$$

case (ii) $$n \log\left[\frac{r_n^{(2)}\{n-r_n^{(1)}\}}{r_n^{(1)}\{n-r_n^{(2)}\}}\right]-\tfrac{1}{2}n(\theta_1+\theta_2), \qquad \dots \quad (4.4)$$

and the stopping limits are given by (2.12) with $I^{\theta\theta}$ equal to

case (i) $$\frac{r_n^{(2)}\{n-r_n^{(2)}\}+r_n^{(1)}\{n-r_n^{(1)}\}}{n^2}\,; \qquad \dots \quad (4.5)$$

case (ii) $$\frac{n^2}{r_n^{(2)}\{n-r_n^{(2)}\}}+\frac{n^2}{r_n^{(1)}\{n-r_n^{(1)}\}}\,. \qquad \dots \quad (4.6)$$

To illustrate these results, consider the expected numbers of pairs to reach a decision, in comparable situations, of three tests :

(A) Wald's test for the odds ratio;

(B) the maximum likelihood test (4.3) for the difference of probabilities;

(C) the maximum likelihood test (4.4) for the difference of log odds.

Two situations will be considered, one in which the proportion of pairs contributing to (A) is near its maximum of $\frac{1}{2}$, the other in which the proportion is much lower. The details are

$$\text{I} \quad (a) \quad \lambda_1 = \lambda_2 = 0.5 \quad \text{versus} \quad (b) \quad \lambda_1 = 0.5, \lambda_2 = 0.6;$$

$$\text{II} \quad (a) \quad \lambda_1 = \lambda_2 = 0.1 \quad \text{versus} \quad (b) \quad \lambda_1 = 0.1, \lambda_2 = 0.2.$$

For test (A) the base hypotheses are of odds ratios I : 1 versus 1.5; II : 1 versus 2.25. For test (B) the base hypotheses are of differences of 0 and 0.1, in both cases. For test (C) the base hypotheses are of differences I : 0 versus log 1.5; II : 0 versus log 2.25.

The comparison of (A), (B) and (C) does not depend on the particular α and β chosen, but for definiteness we take $\alpha = \beta = 0.05$. Note that if all sample proportions are fairly near $\frac{1}{2}$, expressions (4.5) and (4.6) are near to $\frac{1}{2}$ and 8 respectively.

In both cases the estimate of ϕ is correlated with $\hat{\theta}$, so that, for given α, β, there is an increase in mean sample size as compared with the situation in which ϕ is known.

TABLE. MEAN NUMBER OF PAIRS REQUIRED IN SEQUENTIAL TESTS (A), (B), (C) UNDER HYPOTHESES I(a) $\lambda_1 = \lambda_2 = 0.5$, (b) $\lambda_1 = 0.5, \lambda_2 = 0.6$; II(a) $\lambda_1 = \lambda_2 = 0.1$, (b) $\lambda_1 = 0.1, \lambda_2 = 0.2$

test	asymptotic mean sample size			
	case I		case II	
	hyp. (a)	hyp. (b)	hyp. (a)	hyp. (b)
(A) Wald : proportion of "effective" pairs	260 0.5	263 0.5	184 0.18	134 0.26
(B) m.l. on difference of probabilities	265	260	95	132
(C) m.l. on log odds ratio	258	263	179	140

More extensive calculations are desirable. The general conclusion seems to be that when the proportion of "effective" pairs in Wald's test is not too small, then there is little to choose between the tests, but that, at least in some cases, an appreciable reduction in mean sample size can be obtained from the test (4.3). In interpreting the Table, recall that the results are asymptotic so that it is unlikely that the units figure has much meaning.

One further point concerns the increase, if any, in sample size due to the introduction of a nuisance parameter. This can be assessed by comparing the variances of the maximum likelihood estimators of the relevant parameters. Thus, suppose we compared the hypotheses $\lambda_2 = 0.5$, $\lambda_2 = 0.6$ and wanted an operating characteristic depending only on λ_2. Then only single observations, not pairs, need be made. Further, the variance of λ_2 is approximately $\{2\sqrt{n}\}^{-1}$. But in test (B) the variance of θ after n pairs is approximately $\{\sqrt{(2n)}\}^{-1}$, because the difference of two proportions is taken. Thus in effect there is a fourfold increase in sample size due to the introduction of the nuisance parameter. On the other hand, suppose that $\lambda_1 = \phi - \frac{1}{2}\theta$, $\lambda_2 = \phi + \frac{1}{2}\theta$ and $\phi = \frac{1}{2}$. Then $\hat{\phi}$, $\hat{\theta}$ are uncorrelated and the mean sample size is the same whether or not ϕ is regarded as known.

5. Discussion

Two main extensions of the results of this paper are worth considering. One is to processes observed in continuous time. For example, the sequential comparison of survivor curves (Armitage, 1959), using maximum likelihood estimators recomputed continuously in time, would require such an extension. The second development, which is likely to be much more difficult to accomplish, is to obtain more refined approximations to the properties of the tests. This would involve considering the stochastic process of maximum likelihood estimators corresponding to all possible sample sizes and representing this by something more refined than a sequence of means of independent random variables.

References

Armitage, P. (1957) : Restricted sequential procedures. *Biometrika*, **44**, 9-26.

——— (1959) : The comparison of survivor curves. *J. Roy. Stat. Soc.*, A, **122**, 279-300.

Barnard, G. A. (1946) : Sequential tests in industrial statistics. *Suppl. J. Roy. Stat. Soc.*, **8**, 1-26.

——— (1952) : The frequency justification of certain sequential tests. *Biometrika*, **39**, 144-150.

Bartlett, M. S. (1946a) : The large sample theory of sequential tests. *Proc. Camb. Phil. Soc.*, **42**, 239-244.

——— (1946b) : Discussion of paper by G. A. Barnard. *Suppl. J. Roy. Stat. Soc.*, **8**, 22-23.

Cox, D. R. (1952) : Sequential tests for composite hypotheses. *Proc. Camb. Phil. Soc.*, **48**, 290-299.

——— (1958) : Two further applications of a model for binary regression. *Biometrika*, **45**, 562-565.

Girshick, M. S. (1946) : Contributions to the theory of sequential analysis, I. *Ann. Math. Stat.*, **17**, 123-143.

National Bureau of Standards (1951) : *Tables to Facilitate Sequential t Tests.* U.S. Department of Commerce, Washington.

Rushton, S. (1950) : On a sequential *t* test. *Biometrika*, **37**, 326-333.

Wald, A. (1947) : *Sequential Analysis*, Wiley, New York.

Paper received : October, 1962.

ON ASYMPTOTIC EXPANSIONS FOR SUMS OF INDEPENDENT RANDOM VARIABLES WITH A LIMITING STABLE DISTRIBUTION

By HARALD CRAMÉR

University of Stockholm

SUMMARY. The sum of n independent and identically distributed random variables will, under certain known conditions and after appropriate normalization, converge in distribution. The limiting distribution will then belong to the class of stable distributions. In the particular case when the limiting distribution is normal it is well known that, under certain additional conditions, the distribution function $F_n(x)$ of the normalized sum admits an asymptotic expansion in powers of $n^{-1/2}$, valid as n tends to infinity. The present paper considers the analogous question for the case of a non-normal stable limiting distribution. Sufficient conditions for the validity of an asymptotic expansion for $F_n(x)$ in this case are given, and explicit expressions for the terms of the expansion are deduced. As in the normal case, the terms are functions of x, multiplied with certain negative powers of n.

1. Introduction

Let $x_1, x_2, \ldots,$ be independent and identically distributed random variables, with a given common distribution function (d.f.) $F(x)$, and the characteristic function (c.f.)

$$f(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x).$$

If it is possible to find constants $B_n > 0$ and A_n such that the d.f. of the normalized sum

$$\frac{x_1+x_2+\ldots+x_n-A_n}{B_n} \qquad \ldots \quad (1.1)$$

tends to a limiting d.f. $G(x)$ as n tends to infinity, we say that $F(x)$ belongs to the *domain of attraction* of $G(x)$.

Denoting by $F_n(x)$ and $f_n(t)$ the d.f. and c.f. of the random variable (1.1) we then have, for any continuity point x of $G(x)$, and for all real t

$$F_n(x) = F^{n*}(A_n+B_nx) \to G(x),$$

$$f_n(t) = e^{-itA_n/B_n}\left[f\left(\frac{t}{B_n}\right)\right]^n \to g(t), \qquad \ldots \quad (1.2)$$

where $F^{n*}(x)$ denotes the n times repeated convolution of $F(x)$ with itself, while $g(t)$ is the c.f. corresponding to the d.f. $G(x)$.

It is well known that, in order that $G(x)$ should possess a domain of attraction in this sense, it is necessary and sufficient that $G(x)$ should be the d.f. of a *stable* distribution. The c.f. $g(t)$ is then given by the expression

$$\log g(t) = hit - c|t|^{\alpha}[1+\mu i\omega(t, \alpha)], \qquad \ldots \quad (1.3)$$

where h is any real constant, while c, α and μ are constants such that $c \geqslant 0$, $0 < \alpha \leqslant 2$, $-1 \leqslant \mu \leqslant 1$. Finally,

$$\omega(t, \alpha) = \begin{cases} \operatorname{sgn} t \,.\, tg \dfrac{\alpha\pi}{2} & \text{when } \alpha \neq 1, \\ \operatorname{sgn} t \,.\, \dfrac{2}{\pi} \log|t| & \text{when } \alpha = 1. \end{cases} \qquad \ldots \quad (1.4)$$

The case $c = 0$ is trivial, and will be excluded from our considerations, so that we may throughout assume $c > 0$.

When the limiting relations (1.2) hold, we can evidently always, by an appropriate modification of A_n and B_n obtain a limiting distribution such that, in the expression (1.3) for the c.f., we have $h = 0$ and $c = 1$.

The parameter α is called the *characteristic exponent* of the stable distribution. For $\alpha = 2$ the distribution is normal; for $0 < \alpha < 2$ we have non-normal stable distributions.

In order that $F(x)$ should belong to the domain of attraction of a non-normal stable law with the characteristic exponent α, it is necessary and sufficient that we should have, as $x \to +\infty$,

$$F(-x) \sim \frac{h_1(x)}{x^\alpha}, \qquad 1 - F(x) \sim \frac{h_2(x)}{x^\alpha}, \qquad \ldots \quad (1.5)$$

where for any constant $k > 0$ $\quad \dfrac{h_i(kx)}{h_i(x)} \to 1, \qquad (i = 1, 2),$

while the ratio $h_1(x)/h_2(x)$ tends to a constant limit (cf. Gnedenko and Kolmogorov (1949, p. 175); also Dynkin (1955)). We may, e.g., take $h_i(x) = c_i(\log x)^q$.

Particularly important is the case when the $h_i(x)$ can be taken as constants, so that

$$F(-x) \sim \frac{c_1}{x^\alpha}, \qquad 1 - F(x) \sim \frac{c_2}{x^\alpha}, \qquad \ldots \quad (1.6)$$

where $c_1 \geqslant 0$, $c_2 \geqslant 0$, and $c_1 + c_2 > 0$. The normalizing constants B_n in (1.1) can then be determined by the formula

$$B_n = bn^{1/\alpha}, \qquad \ldots \quad (1.7)$$

where $b > 0$ is a constant. In this case, $F(x)$ is said to belong to the *domain of normal attraction* of the limiting stable law (cf. Gnedenko and Kolmogorov (1949) p. 191).

In two papers (Cramér, 1925, 1928) and in my Cambridge Tract (Cramér, 1937) I proved the existence of an asymptotic expansion for $F_n(x)$, as $n \to \infty$ in the

case of a *normal* limiting distribution. I also considered the case when the x_i in (1.1) are not identically distributed. Moreover, I gave (Cramér, 1928) a similar asymptotic expansion for the probability density $F'_n(x)$.

All these results, for the normal limit law, have since been generalized and improved by various authors [Esseen (1944); Gnedenko and Kolmogorov (1949); Petrov (1959) and others]. For the case of a non-normal stable limit law, on the other hand, comparatively little seems to be known. Some results have been given in two papers by Bergström (1951 and 1953), and I have recently published a preliminary note on the subject in the Anniversary Volume dedicated to G. Pólya (Cramér, 1962).

In the present paper, the problem of asymptotic expansions for $F_n(x)$ will be considered for the case of normal attraction to a non-normal stable law. It will be shown that, when appropriate conditions are imposed on the given d.f. $F(x)$, there exists for the d.f. $F_n(x)$ given by (1.2) an asymptotic expansion as $n \to \infty$, the successive terms of which tend to zero as certain, in general fractional, powers of $1/n$. Similar expansions can be shown to exist in the case of non-identically distributed variables and also, under appropriate conditions, for the probability density $F'_n(x)$.

In the case of non-normal attraction, the situation is in general quite different. When e.g., the functions $h_i(x)$ in (1.5) are of the form $c_i(\log x)^q$ with a non-integral q, there is, under similar conditions as in the normal attraction case, an asymptotic expansion for $F_n(x)$. However, the successive terms of this expansion only tend to zero as powers of $\dfrac{1}{\log n}$. This case will not be further dealt with in the present paper.

2. Theorems on asymptotic expansions

Throughout the following, α, β and γ will be real constants satisfying the relations

$$0 < \alpha < 2, \qquad \alpha < \beta < \gamma. \qquad \dots \quad (2.1)$$

Further, c_1 and c_2 are non-negative constants, at least one of which differs from zero, while d_1 and d_2 are any real constants. In particular, we may have $d_1 = d_2 = 0$. However, if c_1 or c_2 is equal to zero, we assume that the corresponding d is non-negative.

Let $F(x)$ be a given d.f. satisfying the following conditions (A) and (B) :

(A) *As $x \to +\infty$, we have*

$$F(-x) = \frac{c_1}{x^\alpha} + \frac{d_1}{x^\beta} + r_1(x),$$

$$1-F(x) = \frac{c_2}{x^\alpha} + \frac{d_2}{x^\beta} + r_2(x)$$

$$r_i(x) = O\left(\frac{1}{x^\gamma}\right), \qquad (i = 1, 2).$$

(B) *In the case when $0 < \gamma \leqslant 1$, the functions*

$$r_1(x) \pm r_2(x)$$

are assumed to be monotone for all sufficiently large $x > 0$.

It will be seen that condition (A) implies considerably stronger assumptions about the infinitary behaviour of $F(x)$ than do the relations (1.6). Some assumptions of this kind will be necessary in order to get more precise information concerning the asymptotic behaviour of $F_n(x)$ for large n than the one provided by the limit relation (1.2). However, the method used below will be applicable even if condition (A) is modified in various ways, e.g. by the introduction of further terms in the asymptotic expressions of $F(-x)$ and $1-F(x)$.

For our Theorem 2, we shall also use the following condition (C), bearing on the c.f. $f(t)$ corresponding to $F(x)$:

(C) $$\limsup |f(t)| < 1 \quad \text{as } |t| \to \infty.$$

It is well known that, in particular, this condition is always satisfied when the d.f. $F(x)$ contains an absolutely continuous component.

We shall denote by $G_\alpha(x)$ the stable d.f. corresponding to the c.f. $g_\alpha(t)$ obtained from (1.3) by taking $h = 0$ and $c = 1$, so that

$$g_\alpha(t) = \exp\left[-|t|^\alpha(1+\mu i\omega(t, \alpha)\right], \qquad \ldots \quad (2.2)$$

where we now take
$$\mu = \frac{c_1-c_2}{c_1+c_2}$$
while $\omega(t, \alpha)$ is given by (1.4).

Further, we shall use the notation

$$k = k_1\alpha+k_2(\beta-\alpha)+k_3(2-\alpha)+k_4, \qquad \ldots \quad (2.3)$$

where k_1, k_2, k_3 and k_4 are non-negative integers, at least one of which differs from zero. It follows, in particular, that we always have $k > 0$.

If $P_k(t)$ is a polynomial in t of degree $k_1+k_2+k_3+k_4-1$, with complex coefficients depending on the k_i, we write

$$G_\alpha(x; P_k) = -\frac{1}{\pi}\, \mathscr{I}_m\left[\int_0^\infty t^{k+\alpha-1}P_k(t^\alpha)g_\alpha(t)e^{-itx}dt\right]. \qquad \ldots \quad (2.4)$$

It will be shown below in Lemma 1 that $G_\alpha(x; P_k)$ is a function of the real variable x which is everywhere continuous, has derivatives of all orders, tends to zero as $x \to \pm\infty$, and is of bounded variation over the whole real axis.

We shall now state our two main theorems.

Theorem 1 : *Suppose that α, β and γ are not integers. If the given d.f. $F(x)$ satisfies conditions* (A) *and* (B) *it is possible to choose the normalizing constants A_n and $B_n = bn^{1/\alpha}$ in* (1.1), *and the polynomials P_k such that, as $n \to \infty$,*

$$F_n(x) = G_\alpha(x)+\sum_{0<k<\lambda} G_\alpha(x; P_k)n^{-k/\alpha}+O(n^{-\lambda/\alpha}), \qquad \ldots \quad (2.5)$$

uniformly for all real x, where $\lambda = \min(1, \gamma-\alpha)$. The summation is extended over all k given by (2.3), *which satisfy the inequality $0 < k < \lambda$.*

Theorem 2 : *Suppose that α, β and γ are not integers. If $\gamma-\alpha > 1$, and the given d.f. $F(x)$ satisfies conditions* (A) *and* (C) *the summation in the second member of* (2.5) *may be extended over all k such that $0 < k < \gamma-\alpha$, and the remainder term will be of the order*

$$O(n^{-(\gamma-\alpha)/\alpha}).$$

Remark (1) : If α or β (or both) is an integer, powers of $\log t$ will occur under the integral sign in (2.4), and the terms of the asymptotic expansion in (2.5) will be multiplied by polynomials in $\log n$. If γ is an integer, the majorants of the remainder terms in Theorems 1 and 2 must be multiplied by $\log n$. Since the explicit formulas for these cases are somewhat cumbersome, we shall restrict ourselves to this remark.

Remark (2) : Note that, in the case of Theorem 2, we necessarily have $\gamma > 1$, so that condition (B) does not apply.

Remark (3) : In particular cases, some of the $G_\alpha(x; P_k)$ may be identically zero. An extreme example of this phenomenon is obtained by taking $F(x) = G_\alpha(x)$. For appropriately chosen normalizing constants, $F_n(x)$ will then be identical with $G_\alpha(x)$.

3. Some lemmas

Before we can proceed to the proofs of the main theorems, we shall have to state and prove a number of lemmas.

Lemma 1 : *The function $G_\alpha(x; P_k)$ defined by (2.4) is an everywhere continuous real-valued function of the real variable x, which has derivatives of all orders, tends to zero as $x \to \pm\infty$, and is of bounded variation over the whole real axis. Moreover, we have for $t > 0$*

$$\int_{-\infty}^{\infty} e^{itx} d_x G_\alpha(x;\ P_k) = t^{k+\alpha} P_k(t^\alpha)\, g_\alpha(t). \qquad \ldots \quad (3.1)$$

For $t < 0$, the integral evidently takes the complex conjugate of its value for $-t > 0$.

It will be seen from the expression (2.2) of $g_\alpha(t)$ that the integral in the second member of (2.4), as well as all the integrals obtained by repeated differentiation with respect to x, are absolutely and uniformly convergent for all real x, and tend to zero as $x \to \pm\infty$. It thus only remains to show that $G_\alpha(x; P_k)$ is of bounded variation, and that we have the relation (3.1).

In order to prove the bounded variation property, it is sufficient to show that, for any $q > \alpha$, the integral

$$J(x) = \int_0^\infty t^{q-1} g_\alpha(t) e^{-itx}\, dt$$

is of bounded variation over $(-\infty, \infty)$. The derivative $J'(x)$ can be calculated by formal differentiation under the integral sign, and we have to show that $J'(x)$ is absolutely integrable over $(-\infty, \infty)$. The proof will be different according as $q \leqslant 1$, or $q > 1$.

If $q \leqslant 1$, we must have $0 < \alpha < 1$. For $x > 0$ we then obtain

$$J'(x) = -i \int_0^\infty t^q g_\alpha(t) e^{-itx} dt = -i\, x^{-(q+1)} \int_0^\infty t^q \exp\left[-t^\alpha x^{-\alpha}\left(1 + \mu i t g \frac{\alpha\pi}{2}\right) - it\right] dt.$$

A method used in a similar case by Skorohod (1954) can then be applied to show that the last integral is bounded as $x \to +\infty$. Since $|J'(-x)| = |J'(x)|$, we thus have

$$J'(x) = O(|x|^{-(q+1)})$$

as $|x| \to \infty$, which shows that $|J'(x)|$ is integrable.

For $q > 1$, a repeated partial integration shows that

$$J'(x) = -i \int_0^\infty H(t, x) g_\alpha''(t) dt,$$

where
$$H(t, x) = \int_0^t du \int_0^u v^q e^{-ivx} dv.$$

By elementary calculations we obtain

$$|H(t, x)| < K \frac{t^q}{x^2},$$

and it then follows from the expression (2.2) for $g_\alpha(t)$ that in this case $J'(x) = O(x^{-2})$ so that $J'(x)$ is integrable.

Thus $G_\alpha(x; P_k)$ is of bounded variation over the whole real x-axis, so that the Stieltjes integral in (3.1) exists. It is then easily seen that (3.1) is the Fourier-Stieltjes transform which is reciprocal to (2.4). Since both integrals are absolutely convergent, (3.1) is certainly valid, and the proof of Lemma 1 is completed.

We now proceed to two lemmas concerning the properties of the c.f. $f_n(t)$ defined by (1.2), for large values of n.

Lemma 2 : *Suppose that α, β and γ are not integers, and that $F(x)$ satisfies conditions* (A) *and* (B). *Then it is possible to choose the normalizing constants A_n and $B_n = bn^{1/\alpha}$ in* (1.1), *and the polynomials P_k such that, uniformly for $0 < t \leqslant n^{1/\alpha - 1/3}$, we have*

$$f_n(t) = e^{-Ct^\alpha} [1 + \sum_{0<k<\gamma-\alpha} t^{\alpha+k} P_k(t^\alpha) n^{-k/\alpha}] + O[t^\gamma(1+t^\delta) e^{-\frac{1}{2}t^\alpha} n^{-(\gamma-\alpha)/\alpha}],$$

where
$$C = 1 + \frac{c_1 - c_2}{c_1 + c_2} \, itg \, \frac{\alpha\pi}{2}, \qquad \dots \quad (3.2)$$

while δ is a positive constant only dependent on α, β and γ. The summation index k is given by (2.3), *and the summation is extended over all non-negative integers $k_1, \dots, k_4$ such that $0 < k < \gamma - \alpha$. For $-n^{1/\alpha - 1/3} \leqslant t < 0$ we have, of course, $f_n(t) = \overline{f_n(-t)}$.*

In order to prove this lemma, we shall first deduce an expansion of the c.f. $f(t)$ for small positive values of t with an error term of the order $O(t^\gamma)$. We use in the sequel the notation $Q(it)$ to denote an unspecified polynomial in the argument it, with real coefficients, of degree $< \gamma - 1$, not necessarily the same in different formulas. If $\gamma < 1$, the polynomial Q will be identically zero.

By partial integration we obtain

$$f(t) = 1 - t \int_0^\infty (1 - F(x) + F(-x)) \sin tx \, dx + it \int_0^\infty (1 - F(x) - F(-x)) \cos tx dx.$$

The integrals from 0 to 1 can be developed in convergent power series. In the integrals from 1 to ∞, we replace $F(-x)$ and $1 - F(x)$ by their expressions according to

condition (A) and then obtain, using the known properties of integrals of the form $\int x^{\gamma} \cos x\, dx$ and $\int x^{\gamma} \sin x\, dx$, assuming $t > 0$,

$$f(t) = 1 - Lt^{\alpha} - Mt^{\beta} - R(t) - itQ(it) - O(t^{\gamma}),$$

where
$$L = \left[c_1 + c_2 + i(c_1 - c_2) tg \frac{\alpha\pi}{2} \right] \int_0^{\infty} \frac{\sin x}{x^{\alpha}}\, dx,$$

$$M = \rho(d_1 + d_2) + i\sigma(d_1 - d_2),$$

$$R(t) = t \int_1^{\infty} (r_1(x) + r_2(x)) \sin tx\, dx + it \int_1^{\infty} (r_1(x) - r_2(x)) \cos tx\, dx.$$

Here ρ and σ are real constants only depending on β. We now show that $R(t)$ is, for small $t > 0$, of the form $itQ(it) + O(t^{\gamma})$. This will be explicitly shown for the cosine term; the sine term can be dealt with in the same way.

Consider first the case $0 < \gamma < 1$. Writing $r(x) = r_1(x) - r_2(x)$, we have, denoting by K an unspecified constant,

$$\left| it \int_1^{1/t} r(x) \cos tx\, dx \right| = \left| \int_t^1 r\left(\frac{x}{t}\right) \cos x dx \right| < Kt^{\gamma} \int_0^1 \frac{dx}{x^{\gamma}},$$

and by the second mean value theorem, assuming t so small that $r(x)$ is monotone for $x > 1/t$,

$$\left| it \int_{1/t}^{\infty} r(x) \cos tx\, dx \right| = \left| \int_1^{\infty} r\left(\frac{x}{t}\right) \cos x\, dx \right| = \left| r\left(\frac{1}{t}\right) \int_1^{z} \cos dx \right| < Kt^{\gamma},$$

so that our assertion is proved for this case.

Now suppose $\gamma > 1$, and let $2h-1 < \gamma < 2h+1$, where h is a positive integer. Then

$$it \int_1^{\infty} r(x) \cos tx\, dx = it \int_1^{\infty} r(x) \left(\cos tx - 1 + \dots \pm \frac{(tx)^{2h-2}}{(2h-2)!} \right) dx + itQ(it),$$

where Q is an even polynomial of degree $2h-2 < \gamma - 1$. Majorating the cosine difference in two different ways, we find that the first term in the second member is majorated by

$$Kt^{\gamma} \int_t^{\infty} \left| \cos x - 1 + \dots \pm \frac{x^{2h-2}}{(2h-2)!} \right| \frac{dx}{x^{\gamma}} < Kt^{\gamma} \left(\int_0^1 x^{2h-\gamma}\, dx + \int_1^{\infty} x^{2h-2-\gamma}\, dx \right).$$

Since $2h-\gamma > -1$, while $2h-2-\gamma < -1$, both integrals are convergent, and our assertion is proved. We thus finally have, as $t > 0$ tends to zero,

$$f(t) = 1 - Lt^{\alpha} - Mt^{\beta} - itQ(it) - O(t^{\gamma}). \qquad \dots \quad (3.3)$$

We now choose the normalizing constants B_n in (1.1) so that

$$B_n = bn^{1/\alpha},$$

where b is the positive root of the equation

$$b^{\alpha} = (c_1 + c_2) \int_0^{\infty} \frac{\sin x}{x^{\alpha}}\, dx.$$

In (3.3) we replace t by t/B_n, and write as an abbreviation

$$\tau = \frac{t}{n^{1/\alpha}}. \qquad \dots \quad (3.4)$$

It will then be seen that as long as t remains confined to the interval $0 < t \leqslant n^{1/\alpha - 1/3}$, we shall have $0 < \tau \leqslant n^{-1/3}$, so that τ tends to zero as $n \to \infty$, uniformly for all t in the interval considered. Hence we obtain, taking account of the value of the constant L,

$$f\left(\frac{t}{B_n}\right) = f\left(\frac{\tau}{b}\right) = 1 - C\tau^{\alpha} - D\tau^{\beta} - i\tau Q(i\tau) + O(\tau^{\gamma}), \qquad \dots \quad (3.5)$$

where C is given by (3.2) while D, as well as the coefficients of the polynomial Q, is independent of n and t. The error term is $O(\tau^{\gamma})$, uniformly for all t in the interval considered.

We now choose the normalizing constants A_n in (1.1) so that

$$A_n = -nbq,$$

where q is the constant term of the polynomial $Q(i\tau)$ appearing in (3.5). From (1.2) we then obtain

$$f_n(t) = e^{nqi\tau}\left[f\left(\frac{\tau}{b}\right)\right]^n$$

and further by means of (3.5)

$$\log f_n(t) = n \log \left[e^{qi\tau} f\left(\frac{\tau}{b}\right)\right] = n \log\,[1 - Z + O(\tau^{\gamma})],$$

where

$$Z = (C\tau^{\alpha} + D\tau^{\beta})e^{qi\tau} + i\tau Q_0(i\tau), \qquad \dots \quad (3.6)$$

$Q_0(i\tau)$ denoting a polynomial in $i\tau$, with the same properties as before, but *without a constant term*. Since $\alpha < \beta$ and $\alpha < 2$, it follows that $Z = O(\tau^{\alpha})$, so that $Z \to 0$ as $n \to \infty$. Consequently

$$\log f_n(t) = n \log (1 - Z) + O(n\tau^{\gamma}) = -n \sum_{\nu=1}^{p} \frac{1}{\nu} Z^{\nu} + O(n\tau^{\gamma}),$$

the integer p being taken so that $p\alpha < \gamma \leqslant (p+1)\,\alpha$. Expanding the power Z^{ν} by means of (3.6), we obtain a linear aggregate of powers of τ, each term having an exponent of the form

$$j = j_1\alpha + j_2\beta + 2j_3 + j_4 = (j_1 + j_2 + j_3 - 1)\alpha + j_2(\beta - \alpha) + j_3(2 - \alpha) + j_4 + \alpha,$$

where the j_i are non-negative integers such that $j_1 + j_2 + j_3 = \nu$. It will be seen that we always have $j > \alpha$, except only in the case when $\nu = j_1 = 1$, $j_2 = j_3 = j_4 = 0$. Thus we may write, replacing j by $k + \alpha$ and using (3.4),

$$\log f_n(t) = -nC\tau^{\alpha} + n \sum_{k} D_k \tau^{k+\alpha} + O(n\tau^{\gamma}) = -Ct^{\alpha} + t^{\alpha} \sum_{k} D_k \tau^{k} + O(t^{\alpha}\tau^{\gamma-\alpha}), \qquad \dots \quad (3.7)$$

where k is given by

$$k = k_1\alpha+k_2(\beta-\alpha)+k_3(2-\alpha)+k_4, \qquad \dots \quad (3.8)$$

the k_i running through all non-negative integers such that

$$0 < k < \gamma-\alpha, \qquad \text{and} \qquad k_1+1 \geqslant k_2+k_3.$$

(The second inequality is another way of writing the evident inequality $j_1+j_2+j_3 \geqslant j_2+j_3$.) In particular, the powers τ^{α}, $\tau^{\beta-\alpha}$, $\tau^{2-\alpha}$ and τ will always appear in the sum in the last member of (3.7), in so far as their exponents are $<\gamma-\alpha$. This remark will be used in a moment.

Taking

$$\lambda = \min(\alpha, \beta-\alpha, 2-\alpha) > 0,$$

(3.7) may be written

$$\log f_n(t) = -Ct^{\alpha}+U+V,$$

where

$$U = t^{\alpha} \sum_{k} D_k \tau^k = O(t^{\alpha}\tau^{\lambda}),$$

$$V = O(t^{\alpha}\tau^{\gamma-\alpha}).$$

It follows in particular that for all sufficiently large n, and for all t in the interval considered, we have

$$|U| < \frac{1}{4}t^{\alpha} \qquad \text{and} \qquad |V| < \frac{1}{4}t^{\alpha}.$$

Further

$$f_n(t) = e^{-Ct^{\alpha}}.\, e^{u}.\, e^{v} = e^{-Ct^{\alpha}}\left[\sum_{\nu=0}^{q}\frac{u^{\nu}}{\nu!}+O(U^{q+1}e^{1/4t^{\alpha}})\right][1+O(Ve^{1/4t^{\alpha}})].$$

Taking here the integer q such that $q\lambda < \gamma-\alpha \leqslant (q+1)\lambda$, we have, observing that according to (3.2) the real part of C is equal to unity,

$$f_n(t) = e^{-Ct^{\alpha}}\left(1+\sum_{\nu=1}^{q}\frac{u^{\nu}}{\nu!}\right)+O[(t^{\alpha}+t^{(q+1)\alpha})\tau^{\gamma-\alpha}e^{-\frac{1}{2}t^{\alpha}}]. \qquad \dots \quad (3.9)$$

Now we have for $\nu = 1, \dots, q$

$$U^{\nu} = t^{\nu\alpha}[\sum_{k} D_{k\nu}\tau^k+O(\tau^{\gamma-\alpha})] \qquad \dots \quad (3.10)$$

where the $D_{k\nu}$ are constants, while k is given by (3.8) the k_i being now certain non-negative integers satisfying

$$0 < k < \gamma-\alpha. \qquad \dots \quad (3.11)$$

Let $k_1, \dots, k_4$ be any given set of non-negative integers satisfying (3.11). It then easily follows that we must have

$$k_1+k_2+k_3+k_4 \leqslant q,$$

so that a term in $U^{k_1+k_2+k_3+k_4}$ will certainly appear in the sum in the second member of (3.9). According to a previous remark, the expansion (3.10) of this power will include a term in

$$\tau^k = (\tau^{\alpha})^{k_1}(\tau^{\beta-\alpha})^{k_2}(\tau^{2-\alpha})^{k_3}\tau^{k_4}.$$

On the other hand, it is obvious that no U^ν with $\nu > k_1+\ldots+k_4$ can include the power τ^k with the given values of the k_i. It follows that we may write

$$\sum_{\nu=1}^{q} \frac{U^\nu}{\nu\,!} = t^\alpha \sum_k \tau^k P_k(t^\alpha)+O[(t^\alpha+t^{(q+1)\alpha})\tau^{\gamma-\alpha}], \qquad \ldots \quad (3.12)$$

where the sum in the second member is extended over all k given by (3.8) and satisfying (3.11), while $P_k(t^\alpha)$ is a polynomial in t^α of degree $k_1+\ldots+k_4-1$. Finally, we obtain from (3.4), (3.9) and (3.12)

$$f_n(t) = e^{-Ct^\alpha}[1+ \sum_k t^{\alpha+k}P_k(t^\alpha)n^{-k/\alpha}]+O[t^\gamma(1+t^{q\alpha})e^{-\frac{1}{2}t^\alpha}\, n^{-(\gamma-\alpha)/\alpha}\,]$$

uniformly for $0 < t \leqslant n^{1/\alpha-1/3}$, so that Lemma 2 is proved.

Lemma 3 : *If* $F(x)$ *satisfies condition* (A) *there are positive constants* p *and* q *such that*

$$|f_n(t)| < e^{-pn^{1/3}}$$

for

$$n^{1/\alpha-1/3} < |t| \leqslant q\, n^{1/\alpha}.$$

When $F(x)$ satisfies condition (A), it is obvious that the modulus of the c.f. $f(t)$ cannot reduce to a constant. Hence we must have $|f(t)| < 1$ for all $t \neq 0$ in some neighbourhood of the origin. Thus we can find $g > 0$ such that $|f(t)| \leqslant h < 1$ for $g \leqslant |t| \leqslant 2g$. It then follows from Lemma 1 of my Cambridge Tract (Cramér, 1937) that

$$|f(t)| \leqslant 1- \frac{1-h^2}{8g^2}\, t^2 \leqslant \exp\left(-\frac{1-h^2}{8g^2}\, t^2\right)$$

for $|t| \leqslant g$. (In the work quoted, this is proved under the assumption that $|f(t)| \leqslant h < 1$ holds for all $|t| \geqslant g$. The only property used in the proof is, however, that this inequality holds for $g \leqslant |t| \leqslant 2g$.)

Further by (1.2), taking $B_n = bn^{1/\alpha}$,

$$|f_n(t)| = |f\left(\frac{t}{bn^{1/\alpha}}\right)|^n,$$

and thus for $n^{1/\alpha-1/3} < |t| \quad gbn^{1/\alpha}$

$$|f_n(t)| \leqslant \exp\left[-\frac{1-h^2}{8b^2g^2}\, n \left(\frac{t}{n^{1/\alpha}}\right)^2\right] < \exp\left(-\frac{1-h^2}{8b^2g^2}\, n^{1/3}\right).$$

Taking $p = \frac{1-h^2}{8b^2g^2}$, $q = gb$, Lemma 3 is proved.

We finally state the following lemma which is contained in one due to Esseen (1944).

Lemma 4 : *Let* $F(x)$ *be any d.f. with the c.f.* $f(t)$. *Let* $G(x)$ *be a real function of bounded variation over the whole real axis, such that* $G(-\infty) = 0$, $G(+\infty) = 1$, *while the derivative* $G'(x)$ *exists everywhere and satisfies* $|G'(x)| < K$ *for some constant* K. *Write*

$$g(t) = \int_{-\infty}^{\infty} e^{itx}\, dG(x).$$

Suppose that for some positive constants T and ϵ we have

$$\int_{-T}^{T} \left| \frac{f(t)-g(t)}{t} \right| dt = \epsilon.$$

Then there are positive constants A and B independent of T and ϵ such that for all real x

$$|F(x)-G(x)| < A\epsilon + \frac{B}{T}.$$

4. Proofs of the theorems

In Lemma 4, we now replace $F(x)$ by $F_n(x)$, and $G(x)$ by

$$G_\alpha(x) + \sum_k G_\alpha(x;\, P_k) n^{-k/\alpha}, \qquad \ldots \quad (4.1)$$

the sum being extended over all k given by (2.3) or (3.8), and satisfying

$$0 < k < \lambda = \min(1, \gamma-\alpha).$$

Then $f(t)$ will be replaced by $f_n(t)$, while according to Lemma 1 we have to replace $g(t)$ by

$$e^{-Ct^\alpha} [1 + \sum_k t^{\alpha+k} P_k(t^\alpha) n^{-k/\alpha}]$$

for $t > 0$, C being given by (3.2), and by the complex conjugate value of $g(-t)$ for $t < 0$.

Further, $G'(x)$ will exist everywhere, and satisfy the inequality $|G'(x)| < K$, where as before K denotes an unspecified constant, independent of x and n.

From Lemma 2 we then obtain

$$\int_{|t| \leqslant n^{1/\alpha - 1/3}} \left| \frac{f(t)-g(t)}{t} \right| dt < K\, n^{-\lambda/\alpha}, \qquad \ldots \quad (4.2)$$

while Lemma 3 gives, taking account of the behaviour of $g(t)$ for large $|t|$,

$$\int_{n^{1/\alpha - 1/3} < |t| \leqslant q n^{1/\alpha}} \left| \frac{f(t)-g(t)}{t} \right| dt < K \log n\, e^{-pn^{1/3}} < K\, n^{-(\gamma-\alpha)/\alpha}.$$

Thus we may take $T = qn^{1/\alpha}$ and $\epsilon = Kn^{-\lambda/\alpha}$ in Lemma 4, which now yields

$$|F_n(x) - G_\alpha(x) - \sum_k G_\alpha(x;\, P_k) n^{-k/\alpha}| < An^{-\lambda/\alpha} + Bn^{-1/\alpha} < Kn^{-\lambda/\alpha} \qquad \ldots \quad (4.3)$$

so that Theorem 1 is proved.

In order to prove Theorem 2, we extend the summation in (4.1) over all k satisfying

$$0 < k < \gamma - \alpha.$$

By Lemma 2, the second member of (4.2) may then be replaced by an expression of the form $Kn^{-(\gamma-\alpha)/\alpha}$. We further observe that, if $F(x)$ satisfies condition (C) there is a constant $r < 1$ such that for $|t| > qn^{1/\alpha}$ we have

$$|f_n(t)| < r^n$$

and hence

$$\int_{qn^{1/\alpha}<|t|\leqslant n^{(\gamma-\alpha)/\alpha}} \left|\frac{f(t)-g(t)}{t}\right| dt < K(r^n \log n + e^{-n^{1/2}}) < Kn^{-(\gamma-\alpha)/\alpha}.$$

Thus in this case we may take $T = n^{(\gamma-\alpha)/\alpha}$ and $\epsilon = Kn^{-(\gamma-\alpha)/\alpha}$ in Lemma 4. The last member of (4.3) will then be replaced by $Kn^{-(\gamma-\alpha)/\alpha}$, and the proof of Theorem 2 is completed.

REFERENCES

BERGSTRÖM, H. (1951) : On asymptotic expansions of probability functions. *Skandinavisk Aktuarieti-dskrift*, 1-33.

——— (1953) : On distribution functions with a limiting stable distribution function. *Arkiv för Matematik*, **2**, Number 25.

CRAMÉR, H. (1925) : On some classes of series used in mathematical statistics. Sixth Scandinavian Mathematical Congress, Copenhaguen, 399-425.

——— (1928) : On the composition of elementary errors. *Skandinavisk Aktuarietidskrift*, 13-74 and 141-180.

——— (1937) : Random variables and probability distributions. *Cambridge Tracts in Mathematics*, No. 36, Cambridge, second edition 1962.

———(1962) : On the approximation to a stable probability distribution. *Essays in honor of George Pólya*, Stanford University Press.

DYNKIN, E. B. (1955) : Some limit theorems for sums of independent random variables with infinite mathematical expectations. *Izvestija Akad. Nauk SSSR, Ser. Math.*, **19**, 247-266. *Selected Translation in Math. Stat. and Prob.*, **1** (1961), 171-190.

ESSEEN, C. G. (1944) : Fourier analysis of distribution functions. *Acta Mathematica*, **77**, 1-125.

GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954) : *Limit distributions for sums of independent random variables.* Cambridge, Mass., Translation from the Russian original, published in 1949.

PETROV, V. V. (1959) : Asymptotic expansions for distributions of sums of independent random variables. *Teorija Verojatnost. i Primenen*, **4**, 220-224.

SKOROHOD, A. V. (1954) : On a theorem concerning stable distributions. *Uspehi Mat. Nauk* (N.S.), **9**, 2, 189-190. *Selected Translations in Math. Stat. and Prob.*, **1** (1961), 169-170.

Paper received : November, 1962.

A SEQUENTIAL PROCEDURE FOR THE BEST POPULATION

By IRWIN GUTTMAN

McGill University and University of Wisconsin

SUMMARY. This paper deals with the problem of choosing the best population from a collection of k given populations, where the term "best" is defined according to a specific criterion which is of interest to the statistician (e. g., largest mean, smallest variance, etc.). A procedure which is of a sequential nature is proposed and its properties studied.

1. Introduction

A pseudo-sequential procedure is given for determining the "best" population in a group of k given populations, with overall confidence β. The procedure is sequential in that after each stage a decision is made as to whether sampling is to be continued or not. The procedure is not sequential in the classical statistical sense in that observations taken in previous stages are not used in subsequent stages.

Each stage requires the use of a selection procedure of the Gupta-Sobel kind (see Section 2), with varying values at each stage for the value of their confidence coefficient P^*. The procedure terminates when either one population is left or the condition of the stopping rules of Section 2 are met. A condition is given for the procedure to terminate with probability one.

2. Formulation of the procedure

We are given a collection π of k populations $\pi_1, \pi_2, \ldots, \pi_i, \ldots, \pi_k$ in which there exists a so-called "best" population, where the population is defined to be "best" according to a specific definition of the following kind.

Suppose the π_i are distributed with probability density function $f(x; \theta_i)$, where possible θ is vector-valued. Consider a real-valued function of θ, $h = g(\theta)$, where g is known. Then the "best" population is that population which has largest value among the $h_i = g(\theta_i)$, $i = 1, \ldots, k$. We assume there is an ordering of the $(h_1, \ldots, h_k)$ into

$$h_{[1]} < h_{[2]} < \ldots < h_{[k-1]} < h_{[k]}. \qquad \ldots \quad (2.1)$$

The sequential procedure proposed below is based on a non-sequential Gupta-Sobel type procedure, which it is now necessary to describe (for examples see Gupta and Sobel, 1960 and Guttman, 1961). A sample of size n independent observations is drawn independently from each population and an "appropriate" selection statistic S_i, $i = 1, \ldots, k$ is computed. (S is appropriate in the sense that $E(S)$ is either $g(\theta)$ or a monotone function of $g(\theta)$). A rule is then formulated of the following type :

$$\text{Retain population} \quad \pi_i \quad \text{if} \quad S_i \epsilon \omega_{n,k}(P^*, \mathbf{S}) \qquad \ldots \quad (2.2)$$

where

(i) $\omega_{n,k}(P^*, \mathbf{S})$ is a random linear set contained in the sample space of S_i and depends on $\mathbf{S} = (S_1, \ldots, S_k)$; it is such that the probability that the "best" population is retained (if this event occurs, it is called a correct selection, and

denoted by CS) is at least P^*. In other words, the probability that the value of S computed from the best population falls in $\omega_{n,k}$ (P^*, $\mathbf{S}$) is at least P^*, and

(ii) it is assumed that the

$$\inf_{\text{all } \theta_i} P(\text{CS}; \theta_1, \ldots, \theta_k) = P^*, \quad \ldots \quad (2.3)$$

where

$$0 < P^* < 1.$$

In general, this procedure leaves us with a retained subset of the k populations, and the size of the retained subset could be 1 or 2 or ... or k, and in general need not be one.

In fact if we let Y_i denote a chance variable which equals 1 if π_i is included in the selected subset, and zero otherwise, then we have that

E(size of the retained subset)

$$= E\{\sum_{i=1}^{k} Y_i\} = \sum_{i=1}^{k} P(\pi_i) \text{ is selected}$$

$$= \sum_{i=1}^{k} P(S_i \epsilon \omega_{n,k} (P^*, \mathbf{S})). \quad \ldots \quad (2.4)$$

Now the above type of Gupta-Sobel (GS) procedure leaves open a legitimate question, viz, if there are two or more populations in the retained subset, which is the best one? The probability that the "best" one is in the retained subset, is of course given in (2.2), i.e., $\Pr(\text{CS}) \geqslant P^*$.

To answer this question we now formulate a sequential procedure (SP). We index the stages of the SP by t. Essentially, we use a GS selection rule of the type (2.2) at each stage, and it is interesting to note then, that if a GS selection rule exists, then a sequential procedure of the type described below automatically exists. The stopping rules are as follows.

(A) At stage t, use a GS selection rule of type (2.2) with $P^* = P_t^*$ given by

$$P_t^* = 1 - \frac{1-\beta}{2^t} \quad \ldots \quad (2.5)$$

and draw independent samples of size n_t independent observations from each of the populations retained.

(B) Let k_t be the number of populations retained. If the number of populations at stage t', say k_t, is greater than one, continue the procedure if and only if $t' < t_0$, where t_0 is defined below, and if $k_t = 1$ for some $t < t_0$, stop the procedure.

(C) If M units of capital are available to spend on this procedure, and if at stage t_0, $k_{t_0} > 1$, then stop the procedure, where t_0 is the largest integer for which

$$\sum_{i=1}^{t_0} k_t n_t d \leqslant M \quad \ldots \quad (2.6)$$

where d is the cost per observation.

Now, let CS denote the event that when the above SP terminates that we have the best population; then if the procedure terminates because of (B), CS is the event that this one population is the best one, while if the procedure is terminated because of the economic condition (C) then we are left with a subset of π and CS is the event that the best population is in the retained subset. We now state the following :

Theorem 2.1 : *When the procedure* SP *given by* (A)-(C) *above is terminated, say at stage* t_r, *then the* Pr(CS) *is given by*

$$\Pr(\mathrm{CS}) \geqslant \beta.$$

Proof : Let A_t be the event of a correct selection at stage t. Suppose the process terminates at stage t_r. Now, the probability that when the procedure stops, the event CS has occurred is the probability of the occurrence of the event

$$A_1 \cap A_2 \cap \ldots \cap A_{t_r}. \qquad \ldots \quad (2.7)$$

But we have that the

$$Pr(A_1 \cap \ldots \cap A_{t_r}) = 1 - Pr\Big(\bigcup_{t=1}^{t_r} \bar{A}_t\Big) \qquad \ldots \quad (2.8)$$

where $\bar{A}_t$ is the event that at stage t we do not have a correct selection, i.e. the complement of the event A_t, hence we have

$$Pr(A_1 \cap \ldots \cap A_{t_r}) \geqslant 1 - \sum_{t=1}^{t_r} P(\bar{A}_t). \qquad \ldots \quad (2.9)$$

Now from (2.3), (2.5) and part (A) of the SP, we have that $P(\bar{A}_t) \leqslant \dfrac{1-\beta}{2^t}$,

and hence

$$Pr(A_1 \cap \ldots \cap A_{t_r}) \geqslant 1 - (1-\beta) \sum_{t=1}^{t_r} \frac{1}{2^t}. \qquad \ldots \quad (2.10)$$

But

$$1 - (1-\beta) \sum_{t=1}^{t_r} \frac{1}{2^t} \geqslant 1 - (1-\beta) \sum_{t=1}^{\infty} \frac{1}{2^t},$$

and so

$$Pr(A_1 \cap \ldots \cap A_{t_r}) \geqslant \beta. \qquad \ldots \quad (2.11)$$

This completes the proof.

3. A CONDITION FOR TERMINATION WITH PROBABILITY ONE

In this section we assume that there is infinite capital available and thus part (C) of the stopping rule of the SP may be ignored. It is interesting to note that Theorem 2.1 still holds for if $t_r = \infty$, (2.11) still holds.

Now, we will say that the SP is in state γ if at any stage t of the procedure there are exactly γ populations retained, where $\gamma = k, \ldots, 1$. From the very nature of the SP, it is clear that the states form a Markov Chain with non-stationary transition probabilities

$$P(k_{t+1} = \alpha \mid k_t = \gamma). \qquad \ldots \quad (3.1)$$

We denote these transition probabilities by $p_{\gamma\alpha}(t, t+1)$ and note that these probabilities are dependent on $\omega_{n_t, \gamma}(P_t^*, \mathbf{S}_t)$ and that $1 \leqslant \alpha \leqslant \gamma = k_t \leqslant k$. Note that $p_{\gamma\alpha}(t) = 0$ if $\gamma < \alpha$ and that $\sum_{\alpha=1}^{\gamma} p_{\gamma\alpha} = 1$. We may now state

Theorem 3.1 : *Consider the Markov Chain with the above structure. Let* $p_{\alpha\alpha}(t) = 1-\delta_\alpha(t)$, $0 < \delta_\alpha(t) < 1$, for $\alpha \neq 1$. *Then the Markov Chain is absorbed at state* 1 (*i.e. the* SP *terminates at a finite stage*) *if and only if* $\sum\limits_{t=1}^{\infty} \delta_\alpha(t)$ *diverges, all* $\alpha \neq 1$.

Proof: Let $\alpha \neq 1$. The probability that the procedure remains in state α for at least l stages is clearly

$$p_{\alpha\alpha}(t_0+1) \dots p_{\alpha\alpha}(t_0+l) \quad \dots \quad (3.2)$$

where the procedure enters state α at stage t_0, and the probability that it stays at state α for an infinite time is

$$\lim_{l\to\infty} p_{\alpha\alpha}(t_0+1) \dots p_{\alpha\alpha}(t_0+l) = \prod_{l=1}^{\infty} (1-\delta_\alpha(t_0+l)) \quad \dots \quad (3.3)$$

and this probability i.e., this infinite product diverges to zero if and only if $\Sigma\, \delta_\alpha(l)$ diverges. That is, if $\sum\limits_{l=1}^{\infty} \delta_\alpha(l)$ diverges, all α, the probability of remaining in α an infinite time is zero, and because of the structure of the Markov Chain, the SP will thus be absorbed at 1 in a finite time.

Note that we may make the assumption $0 < \delta_\alpha(t) < 1$ since $p_{\gamma\alpha}(t) \neq 0$ or 1, all α and γ in view of (2.3).

As a last remark, we wish to point out that if we have some prior knowledge about the possible differences of the $h_{[i]}$, then it might be possible to use this information to find a "reasonable" value for n_t.

For consulting (2.4), we have

$$E(\text{size of the retained subset}) = \sum_{j=1}^{k_t} P(S_j\, \epsilon\, \omega_{n_t,k_t}(P_t^*, \mathbf{S}_t)). \quad \dots \quad (3.4)$$

But the P function is clearly a function of n_t, k_t, P_t^* and the θ_i, which is to say, (3.4) is a function of n_t, k_t, P_t^* and the $h_{[i]}$. In the special cases where this function can be written as a function of n_t, k_t, P_t^* and the differences $h_{[j]}-h_{[i]}$, $i < j$, and since k_t and P_t^* are known, we can set this function equal to one and solve for the unknown n_t.

ACKNOWLEDGEMENT

The author wishes to thank Professor A. Joffe for his enthusiastic interest and encouragement in this problem and for valuable discussions.

REFERENCES

BECKHOFFER, H. E. (1954) : A single k sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Stat.*, 25, 16.

GUPTA, S. S. (1956) : On a decision rule for a problem in ranking means. Report No. 150 ; Institute of Statistics, University of North Carolina.

GUPTA, S. S. and SOBEL, M. (1957) : On a statistic which arises in selection and ranking problems. *Ann. Math. Stat.*, 28, 957.

——— (1960) : Selecting a subset containing the best of several binomial populations. *Contributions to Probability and Statistics, in honour of Harold Hotelling*, Stanford University Press, 224.

GUTTMAN, I. (1961) : Best populations and tolerance regions. *Ann. Stat. Math.*, Tokyo, 13, 9.

Paper received : April, 1962.
Revised : September, 1962.

LONG-CHAIN POLYMERS AND SELF-AVOIDING RANDOM WALKS

By J. M. HAMMERSLEY

Institute of Statistics, University of Oxford

SUMMARY. Self-avoiding random walks provide a simplified model of long-chain polymers. This paper gives some qualitative theorems on the behaviour of such walks. The analysis will be continued in a subsequent paper.

A polymer macromolecule consists of a long main-chain of atoms and some lesser side-assemblies (side-chains, side-rings, etc.) like ribs attached to a flexible backbone. In chemical solution its shape is essentially that of its main-chain, long and wriggly like a random walk with many (10^3 or 10^4) steps. Successive steps are of pretty uniform length and random direction; but these directions are not statistically independent, partly because the side-assemblies restrict the freedom of the main-chain and partly because the main-chain interferes with itself. Essentially, the walk is *self-avoiding,* in as much as the atoms have physical size and no two atoms may occupy the same region of space. This constitutes the so-called *excluded volume problem,* on which scores of research papers have been written. Indeed, the literature merits a summarizing review article every three years or so: see, for example, Wall and Hiller (1954), Hermans (1957), or Casassa (1960), each with their lengthy bibliographies. Despite all this, we still know very little about self-avoiding random walks; for the problem is extremely difficult.

The present paper, like many of its predecessors, deals exclusively with self-avoiding Pólya walks. They afford the simplest possible model of the excluded volume problem : indeed, they are an over-simplification of the chemical reality; but it seems presumptuous to embark on more realistic and more complicated kinds of self-avoiding walk until we have first learnt how to handle these simplest self-avoiding walks.

Consider, then, d-dimensional Euclidean space with coordinates represented vectorially by $\boldsymbol{w} = (x, y, \ldots, z)$. We confine our attention to points of the hyper-cubical lattice in this space : that is to say, a *point* will hereafter mean a point whose coordinates are all (positive, negative, or zero) integers. We say that two points are *neighbours* if they are unit distance apart. We define an *n-stepped Pólya walk* to be an ordered sequence of $n+1$ points such that each pair of successive points are neighbours. We write $\boldsymbol{w}_i = (x_i, y_i, \ldots, z_i)$ with $i = 0, 1, \ldots, n$ for the successive points of such a walk; and $P_n(\boldsymbol{w})$ for the class of distinct n-stepped Pólya walks with a prescribed first point $\boldsymbol{w}_0 = \boldsymbol{w}$. A walk is called self-avoiding if all its points are distinct. We write $F_n(\boldsymbol{w})$ for the self-avoiding subset of $P_n(\boldsymbol{w})$. Clearly $P_n(\boldsymbol{w})$ and $F_n(\boldsymbol{w})$ depend upon $\boldsymbol{w}$ only through a trivial translation; and for most purposes we may work with $P_n = P_n(\mathbf{0})$ and $F_n = F_n(\mathbf{0})$, where $\mathbf{0} = (0, 0, \ldots, 0)$. The number of members

of P_n is clearly $(2d)^n$. We write f_n for the number of members of F_n: there is no known explicit formula for f_n, though clearly we have the trivial result

$$1 \leqslant f_n \leqslant 2d(2d-1)^{n-1}, \qquad \dots \quad (1)$$

with equality on the right when $d = 1$, because no step can be in the opposite direction to the previous step. Individual members of P_n and F_n will, by hypothesis, have equal probabilities, $(2d)^{-n}$ and f_n^{-1} respectively. The distance between the two ends of a walk from F_n is

$$r_n = \sqrt{(x_n^2+y_n^2+\dots+z_n^2)} \qquad \dots \quad (2)$$

and this has a mean-square expectation

$$Q_n = Er_n^2 = dEx_n^2 = d \sum_{x=-n}^{n} x^2 f_n(x)/f_n, \qquad \dots \quad (3)$$

where Q_n is defined by (3) and $f_n(x)$ denotes the number of members of F_n having $x_n = x$. Here, Q_n is a measure of the overall size of a walk; and it is natural to ask how it behaves as a function of n, but nothing rigorous has ever been proved about Q_n beyond the trivial inequality

$$0 < Q_n \leqslant n^2, \qquad \dots \quad (4)$$

with equality on the right when $d = 1$. We may compare (4) with the corresponding familiar result $Q_n = n$ when the walks are drawn from P_n instead of F_n.

The literature contains three main forms of attack on the problem: they are respectively the enumerative, the Monte Carlo, and the analytic attacks.

In the enumerative attack, one enumerates all members of F_n for some given n and hence calculates the corresponding statistics Q_n and the like. However, as we shall see presently,

$$[(2d-1)-\log(2d-1)]^n \leqslant f_n \leqslant 2d(2d-1)^{n-1} \qquad \dots \quad (5)$$

is an improved form of (1) ; and this shows that the population F_n is unmanageably large unless n is very small. In the most important practical case $d = 3$, enumeration has only been taken as far as $n = 11$. Enumeration may be either by pencil-and-paper calculations, such as Sykes (1961) performs, or by electronic computer, as O'Flaherty (1961) and Martin (1962) show. The former alternative, better able to exploit the combinatorial symmetries of the problem, has so far proved superior; but future improvements in programming techniques and machine speed may change this shortly.

The Monte Carlo attack examines a small random sample of F_n. The problem of drawing an unbiassed sample is not as simple as it may appear at first sight; but recent advances in sampling techniques have allowed calculations with n as large as 10^3. In particular, notable work has been done by Wall, Hiller, and Atchison (1955), Wall and Erpenbeck (1959a, 1959b), and Marcer (1961). Monte Carlo work lends some provisional support to the conjecture that Q_n behaves rather like $n^{3/2}$ when $d = 2$, and perhaps like a function increasing a little more rapidly than n (for example, $n \log n$) when $d = 3$.

The analytic attack attempts to handle the problem by pure mathematics. All attempts at precise quantitative results, however, encounter insuperable mathematical difficulties; and authors, invariably forced to make one or another simplifying assumption of more or less doubtful validity, have arrived at a wide variety of mutually conflicting, and hence rather unconvincing, approximations. On the other hand, analysis has yielded a small number of qualitative results, taking the form of existence theorems. For example, Hammersley and Welsh (1962) proved the existence of two constants κ and γ, where κ depends only upon d and is known as the *connective constant*, and where γ is an absolute constant, such that

$$\kappa n \leqslant \log f_n \leqslant \kappa n+\gamma n^{1/2}+\log d; \qquad \ldots \quad (6)$$

and Rennie (1961) proved that $\quad 2d-\log d-c \leqslant e^{\kappa}, \qquad \ldots \quad (7)$

where c is the absolute constant

$$c = 3-\sqrt{2}-\log 2+\sum_{r=3}^{\infty} \frac{\log (r+1)}{r[r-\log (r+1)]}\,. \qquad \ldots \quad (8)$$

The purpose of such existence theorems is to show that certain quantities, in which we are interested, have certain functional forms, prescribed apart from one or two unknown parameters. For example, (6) shows that $\log f_n$ is very nearly a linear function of n when n is large, the slope being an unknown parameter κ. We then combine this kind of information with either an enumerative or a Monte Carlo attack, in the former case to select the appropriate extrapolating formula for attaining to larger values of n, and in the latter case to select the appropriate statistical hypotheses and estimates for fitting to the Monte Carlo samples. This kind of combination of attacks seems quite promising : for example, Sykes (1961) used (6) and his enumeration to estimate that

$$e^{\kappa} = \begin{cases} 2.6390 \pm 0.0005 \text{ when } d = 2 \\ 4.152 \pm 0.003 \text{ when } d = 3. \end{cases} \qquad \ldots \quad (9)$$

The present paper is an analytic attack of this qualitative kind, aiming at existence theorems which will generalize (6) and assist an enumerative or Monte Carlo attack on Q_n. We shall make some progress in this direction; but the results are very much less satisfactory than we might expect. Nevertheless, the results seem to be heading in the right direction; and this paper will serve its purpose if it stimulates further and more comprehensive research.

Specifically, we shall try to investigate how $f_n(x)$ behaves for large values of n. If this investigation could be successfully carried through, it would yield all we need to know about Q_n; for we know that

$$f_n = \sum_{x=-n}^{n} f_n(x), \qquad \ldots \quad (10)$$

and we could then calculate Q_n from (3) and (10). We shall succeed in predicting how $f_n(x)$ behaves when x increases like a multiple of n. Our approach is in fact a suitable generalization to dependent variables of Blackwell and Hodges (1959) treatment of the extreme tail of a convolution. Unfortunately, this is not the most interesting range of

values of x. To predict the behaviour of Q_n, we really need to know how $f_n(x)$ behaves when $x/n \to 0$, and perhaps in particular when x is a multiple of $\sqrt{n}$. This may call for some sort of generalization of the central limit theorem; and we are unable to supply this.

Nevertheless, our tools in handling $f_n(x)$ will be very similar to those used customarily in the central limit theorem. We define

$$H_n(t) = \sum_{x=-n}^{n} e^{xt} f_n(x). \quad \ldots \quad (11)$$

Here t is a real number : it has to be real because we have to use inequalities rather than equations in the analysis. Clearly $H_n(0) = f_n$; so that $H_n(t)/H_n(0)$ is the familiar moment-generating function of the random variable x_n. We also define, for $0 \leqslant \beta \leqslant 1$,

$$H_n(\beta, t) = \sum_{x \geqslant \beta n} e^{xt} f_n(x). \quad \ldots \quad (12)$$

Let F_n^* denote the subclass of walks $(x_i, y_i, \ldots, z_i)$ of F_n such that $0 = x_0 < x_i \leqslant x_n$ for $i = 1, 2, \ldots, n$; and let $f_n^*(x)$ be the number of members of F_n^* which have $x_n = x$. We define $H_n^*(t)$ and $H_n^*(\beta, t)$ by replacing $f_n^*(x)$ in (11) and (12). Of course, we adopt the convention that the number of members of an empty class is zero : thus $f_n^*(x) = 0$ if $x \leqslant 0$.

Theorem 1 :

$$\Sigma_\xi(\xi) f_n^*(x-\xi) \leqslant f_{m+n}^*(x) \leqslant f_{m+n}(x) \leqslant \Sigma_\xi\, f_m(\xi) f_n(x-\xi). \quad \ldots \quad (13)$$

$$H_m^*(\alpha, t) H_n^*(\beta, t) \leqslant H_{m+n}^*\left(\frac{\alpha m+\beta n}{m+n}, t\right) \leqslant H_{m+n}(t) \leqslant H_m(t)\, H_n(t). \quad \ldots \quad (14)$$

$$e^{nt} \leqslant H_n^*(\beta, t) \leqslant H_n(t) \leqslant (2de^{|t|})^n. \quad \ldots \quad (15)$$

Proof of Theorem 1 : Consider a given member of F_m^* with $x_m = \xi$ and a given member of F_n^* with $x_n = x-\xi$. If we translate (bodily without rotation) the latter walk until its first point coincides with the last point of the former walk, we obtain a walk belonging to F_{m+n}^* with $x_{m+n} = x$. Since each distinct pair of members from F_m^* and F_n^* will yield a distinct member of F_{m+n}^*, the left-hand side of (13) follows. The right-hand side follows similarly upon considering any given member of F_{m+n} and translating its last n steps until they belong to F_n, retaining its first m steps as a member of F_m. Distinct members of F_{m+n} will yield distinct pairs of members F_m and F_n. The central inequality of (13) is trivial. From (13) we deduce

$$\sum_{x \geqslant \alpha m} e^{xt} f_m^*(x) \sum_{y \geqslant \beta n} e^{yt} f_n^*(y) \leqslant \sum_{x \geqslant \alpha m} \sum_{y \geqslant \beta n} e^{(x+y)t} f_{m+n}^*(x+y)$$

$$\leqslant \sum_{x+y \geqslant \alpha m+\beta n} e^{(x+y)t} f_{m+n}^*(x+y), \quad \ldots \quad (16)$$

which is equivalent to the left-hand side of (14). Similarly the right-hand side of (14) follows from the right-hand side of (13); and the central inequality of (14) is trivial. The inequalities in (15) follow from (1) and the consideration that the walk $(i, 0, 0, \ldots, 0)$ $(i = 0, 1, \ldots, n)$ contributes to $H_n^*(\beta, t)$ for all $\beta \leqslant 1$. Moreover this is the only walk which contributes when $\beta = 1$.

Theorem 2 : *If A_n is defined for $n = 1, 2, \ldots$ and satisfies*

$$0 < a^{m+n} \leqslant A_{m+n} \leqslant A_m A_n \qquad \ldots \quad (17)$$

for some constant a, then

$$\lim_{n\to\infty} A_n^{1/n} = \inf_{n\geqslant 1} A_n^{1/n}. \qquad \ldots \quad (18)$$

Alternatively, if A_n satisfies

$$0 < A_m A_n \leqslant A_{m+n} \leqslant a^{m+n}, \qquad \ldots \quad (19)$$

then

$$\lim_{n\to\infty} A_n^{1/n} = \sup_{n\geqslant 1} A_n^{1/n}. \qquad \ldots \quad (20)$$

The existence of the limits in (18) *and* (20) *is part of the conclusion of the theorem.*

Proof of Theorem 2 : The fundamental theorem [Hille (1948) Theorem 6.6.1] on subadditive functions states that the inequality

$$B_{m+n} \leqslant B_m + B_n \qquad \ldots \quad (21)$$

implies

$$\lim_{n\to\infty} B_n/n = \inf_{n\geqslant 1} B_n/n, \qquad \ldots \quad (22)$$

where the limit on the left exists, although it may be formally equal to $-\infty$. Theorem 2 now follows upon taking $B_n = \pm\log A_n$.

Theorem 2 is a key theorem in what follows; and, since it is really only a modified form of the fundamental theorem on subadditive functions, it is really the latter theorem that supports the whole analysis. To this extent it is worth mentioning that a generalized version of the fundamental theorem on subadditive functions exists; see Hammersley (1962) for details. This generalized version is not used in the present paper; but it might well prove useful in a more penetrating study of self-avoiding walks.

Theorem 3 : *There exist functions $\theta(t)$ and $\theta(\beta, t)$, both independent of n such that*

$$H_n(t) = \exp\{n[\theta(t) + o_{t,n}]\} \qquad \ldots \quad (23)$$

$$H_n^*(\beta, t) = \exp\{n[\theta(\beta, t) - o_{\beta,t,n}]\} \qquad \ldots \quad (24)$$

where $o_{t,n}$ and $o_{\beta,t,n}$ are non-negative and tend to zero as $n\to\infty$.

Further,

$$t \leqslant \theta(\beta, t) \leqslant \theta(t) \leqslant |t| + \log 2d. \qquad \ldots \quad (25)$$

Proof of Theorem 3 : Put $\alpha = \beta$ in Theorem 1, and apply to Theorem 2 and (15). In fact

$$\theta(t) = \log \inf_{n\geqslant 1} [H_n(t)]^{1/n};\ \theta(\beta, t) = \log \sup_{n\geqslant 1} [H_n^*(\beta, t)]^{1/n}. \qquad \ldots \quad (26)$$

Theorem 4 : *There exists an absolute constant γ such that*

$$H_n(\beta, t) \leqslant d e^{n\theta(\beta,t)+\gamma\sqrt{n}}, \qquad (t \geqslant 0); \qquad \ldots \quad (27)$$

and

$$H_n(\beta, t) \leqslant d e^{n[\beta t+\theta(\beta,0)]+\gamma\sqrt{n}}, \ (t \leqslant 0). \qquad \ldots \quad (28)$$

Proof of Theorem 4 : For a given self-avoiding walk $\{(x_i, y_i, \ldots, z_i)\}_{i=0,i,\ldots,n}$, let ρ be the smallest integer satisfying $x_\rho = \min_{0\leqslant i\leqslant n} x_i$ and let σ be the largest integer satisfying $x_\sigma = \max_{0\leqslant i\leqslant n} x_i$. We call $x_\sigma - x_\rho$ the *extent* of the walk ; and we call $\{(x_i, y_i, \ldots, z_i)\}_{i=0,1,\ldots,\rho}$ and $\{(x_i, y_i, \ldots, z_i)\}_{i=\sigma,\sigma+1,\ldots,n}$ respectively the *lower* and

upper tails of the walk. By a lower (upper) *unfolding* we mean the transformation which reflects the lower (upper) tail in the hyperplane $x = x_l$ (hyperplane $x = x_u$) and which leaves the rest of the walk unchanged. Clearly, unfolding maps an n-stepped self-avoiding walk into an n-stepped self-avoiding walk, and it increases the walk's extent by the extent of the reflected tail.

If we iterate the upper unfoldings of some given walk, and write X_1, X_1+X_2, $X_1+X_2+X_3$, ... for the extent of the original walk and its successive unfoldings, it is easy to see that

$$X_1 \quad X_2 \quad \dots \quad X_n = 0. \qquad \dots \quad (29)$$

In fact X_j will reach zero before (and usually long before) j reaches n; but that need not concern us here. When further iterations yield no fresh walk (i.e. when $X_j = 0$), we may repeat the procedure with lower unfoldings, thus giving walks with successive extents $X_1+X_2+ \dots +X_n+X'_2, \dots, X_1+X_2+ \dots +X_n+X'_2 \dots +X'_n$ where

$$X_1 \geqslant X'_2 \geqslant \dots \geqslant X'_n = 0. \qquad \dots \quad (30)$$

At this stage the walk is fully unfolded on each side, and both its tails have zero extent. Hence, if the walk has become $\{(x^*_i, y^*_i, \dots, z^*_i)\}_{i=0,1,\dots,n}$, we have $X^*_0 \leqslant X^*_i \leqslant X^*_n$. Finally precede this walk by the single point $(x^*_0-1, y^*_0, \dots, z^*_0)$ and then translate this augmented walk (bodily without rotation) until its first point coincides with the origin.

When the original walk is a member of F_n, the above procedure is a mapping of F_n onto F^*_{n+1}. We now show that at least one and at most $\frac{1}{2}e^{\gamma\sqrt{n}}$ distinct members of F_n map in this way into any given member of F^*_{n+1}, where γ is an absolute constant. Let W be the specified walk of F^*_{n+1} and suppose that $X+1$ is its extent. By omitting the first step of W and translating the remainder to start at the origin we obtain one member of F_n which maps into W. On the other hand, let $X_1, X_2, \dots, X_n, X'_2, X'_3, \dots, X'_n$ be any integral solution of

$$X = X_1+X_2+\dots+X_n+X'_2+\dots+X'_n \qquad \dots \quad (31)$$

which also satisfies (29) and (30). We can now fold up the residue of W (after omitting its first step) in a manner intended to restore the originating member of F_n. To be precise, we define an A-left-folding to be a reflection in the hyperplane P_A of all that lies to the left of P_A, where P_A lies a distance A to the right of the left-hand support hyperplane of the walk; and a B-right-folding to be a reflection in P_B of what lies to right of P_B, where P_B lies a distance B to the left of the right-hand support hyperplane of the walk. Then the folding consists of X'_n-left-folding, X'_{n-1}-left-folding, ..., X'_2-left-folding, X_n-right-folding, X_{n-1}-right-folding, ... in that order. Finally, we translate the result to start at the origin. In general, the folded image of a self-avoiding walk is not always self-avoiding. Hence some solutions of (29), (30) and (31) may not yield a member of F_n. However, the subset of all solutions of (29), (30) and (31), which do yield members of F_n, yield all those members of F_n which map into W. Consequently the number of distinct members of F_n which map into W does not exceed the number of solutions of (29), (30) and (31). The number of solutions of (29) and

(31) is $P(X)$, the number of partitions of X. Similarly there are $P(X)$ solutions of (30) and (31). Hence the number of solutions of (29), (30) and (31) does not exceed

$$[P(X)]^2 \leqslant [P(n)]^2 \leqslant \tfrac{1}{2}e^{\gamma\sqrt{n}}, \qquad \dots \quad (32)$$

for some γ because

$$\log P(n) \sim 2\pi\sqrt{(n/3)} \text{ as } n\to\infty \qquad \dots \quad (33)$$

according to Hardy and Ramanujan (1917).

Now let $g_n(x, X)$ denote the number of distinct walks in F_n having $x_n = x$ and also having fully unfolded images of extent $X+1$ in F^*_{n+1}. What we have so far proved amounts to the inequality

$$\sum_x g_n(x, X) \leqslant \tfrac{1}{2}e^{\gamma\sqrt{}}\ f^*_{n+1}(X+1). \qquad \dots \quad (34)$$

The summation on the left of (34) may be restricted to $x \leqslant X$ since $g_n(x, X)$ is zero otherwise. Now, if $t \geqslant 0$

$$\begin{aligned}
H_n(\beta, t) &= \sum_{x\geqslant\beta n} e^{xt} f_n(x) = \sum_{X\geqslant x\geqslant\beta n} e^{xt}\, g_n(x, X) \\
&\leqslant \sum_{X\geqslant x\geqslant\beta n} e^{Xt}\, g_n(x, X) \leqslant \sum_{X\geqslant\beta n} e^{Xt} \sum_{x\leqslant X} g_n(x, X) \\
&\leqslant \sum_{X\geqslant\beta n} e^{Xt}\, \tfrac{1}{2}e^{\gamma\sqrt{n}} f^*_{n+1}(X+1) = \tfrac{1}{2}e^{\gamma\sqrt{n}-t} \sum_{X\geqslant\beta n} e^{(X+1)t} f^*_{n+1}(X+1) \\
&\leqslant \tfrac{1}{2}e^{\gamma\sqrt{n}-t} \sum_{X+1\geqslant\beta(n+1)} e^{(X+1)t} f^*_{n+1}(X+1) \\
&= \tfrac{1}{2}e^{\gamma\sqrt{n}-t}\, H^*_{n+1}(\beta, t) \leqslant \tfrac{1}{2}e^{\gamma\sqrt{n}-t+(n+1)\theta(\beta,t)} \\
&\leqslant \tfrac{1}{2}e^{\gamma\sqrt{n}-t+n\theta(\beta,t)+t+\log 2d} = de^{n\theta(\beta,t)+\gamma\sqrt{n}}. \qquad \dots \quad (35)
\end{aligned}$$

In the derivation of (35) we have used (34), the inequality $0 \leqslant \beta \leqslant 1$, (24), and (25). If $t \leqslant 0$, we have instead

$$H_n(\beta, t) = \sum_{x\geqslant\beta n} e^{xt} f_n(x) \leqslant e^{\beta nt} \sum_{x\geqslant\beta n} f_n(x) = e^{\beta nt} H_n(\beta, 0); \qquad \dots \quad (36)$$

and we can now complete the analysis by applying (35) with $t = 0$.

Theorem 5 : *$\theta(\beta, t)$ is a concave function of β for each fixed t; that is to say*

$$\theta(p\alpha+q\beta, t) \geqslant p\theta(\alpha, t)+q\theta(\beta, t), \quad (0 \leqslant \alpha \leqslant 1,\ 0 \leqslant \beta \leqslant 1) \qquad \dots \quad (37)$$

whenever $p \geqslant 0$, $q \geqslant 0$ and $p+q = 1$. Moreover $\theta(\beta, t)$ is a continuous function of β in $0 < \beta < 1$.

Proof of Theorem 5 : Put $m = n$ in (14), substitute from (24), take logarithms, divide by $2n$, and let $n\to\infty$ to obtain

$$\tfrac{1}{2}\theta(\alpha, t)+\tfrac{1}{2}\theta(\beta, t) \leqslant \theta(\tfrac{1}{2}\alpha+\tfrac{1}{2}\beta, t) \qquad \dots \quad (38)$$

for any fixed α, β. Since (25) shows θ to be bounded, (38) implies the stated results in view of standard theory on convex functions : see Hardy, Littlewood, and Pólya (1934) §3.18.

Theorem 6 :

$$\theta(t) = \theta(0, |t|); \qquad \dots \quad (39)$$

and

$$e^{n\theta(t)} \leqslant H_n(t) \leqslant 2de^{n\theta(t)+\gamma\sqrt{n}} \qquad \dots \quad (40)$$

Proof of Theorem 6 : The definition of $f_n(x)$ shows it to be an even function of x, by symmetry. Hence $H_n(t)$ and $\theta(t)$ are even functions of t, and it is enough to suppose $t \geqslant 0$. Since $H_n^*(0, t) \leqslant H_n(t) \leqslant 2H_n(0, t)$ when $t \geqslant 0$, the result follows from (23), (24), and (27).

Theorem 7 : *For real t, let*

$$K(t) = \log E(e^{X_t}) \qquad \ldots \quad (41)$$

be the cumulant generating function of a bounded random variable X, whose possible values are integers with greatest common divisor 1. *Let b be any number satisfying*

$$EX < b < \sup X, \qquad \ldots \quad (42)$$

and define

$$\mu(b) = \inf_t \ \{K(t)-bt\}. \qquad \ldots \quad (43)$$

Then

$$\text{prob}\left\{ \sum_{i=1}^{N} X_i \geqslant bN \right\} = \exp\{N\mu(b)+o_b(N)\} \text{ as } N \to \infty \qquad \ldots \quad (44)$$

where $X_i (i= 1, 2, \ldots)$ *is a sequence of independent random variables each distributed like* X.

Proof of Theorem 7 : This theorem is simply a considerably weakened form of the main theorem in Blackwell and Hodges (1959), rewritten in notation appropriate to our present needs.

Theorem 8 : *Let* $K_n(t)$ *and* $\mu_n(b)$ *denote the functions* $K(t)$ *and* $\mu(b)$, *which occur in Theorem* 7, *in the particular case when* $X = x_n$, $\{(x_i, y_i, \ldots, z_i)\}_{i=0,1,\ldots,n}$ *being a random member of* F_n. *Then*

$$n\theta(\beta, 0) \leqslant \mu_n(n\,\beta)+ \log f_n, \quad (0 < \beta < 1). \qquad \ldots \quad (45)$$

Proof of Theorem 8 : We have

$$0 = EX < b < \sup X = n; \qquad \ldots \quad (46)$$

so we may take

$$b = \beta n, \quad 0 < \beta < 1. \qquad \ldots \quad (47)$$

The number of Pólya walks, having nN steps, starting from the origin, consisting of N independent pieces, where each piece belongs to F_n (apart from a trivial translation), and finishing to the right of the hyperplane $x = \beta nN$, is the left-hand side of

$$\text{prob}\left\{ \sum_{i=1}^{N} X_i \geqslant \beta nN \right\} f_n^N \geqslant H_{nN}(\beta, 0); \qquad \ldots \quad (48)$$

and the right-hand side of (48) is the number of these walks which also belong to F_{nN}. This proves (48). Now substitute (44) into the left-hand side of (48) and substitute (24) (with $t = 0$) into the right-hand side of the trivial inequality

$$H_{nN}(\beta, 0) \geqslant H_{nN}^*(\beta, 0). \qquad \ldots \quad (49)$$

Take logarithms of the combined results to obtain

$$N\mu_n(\beta n)+o_b(N)+N\log f_n \geqslant nN[\theta(\beta, 0)-o_{\beta,0,nN}]. \qquad \ldots \quad (50)$$

Divide (50) by N and let $N \to \infty$. This proves (45).

Theorem 9 : *$\theta(\beta, 0)$ is a non-increasing continuous concave function of β for $0 \leqslant \beta \leqslant 1$, and it satisfies*

$$0 \leqslant \theta(\beta, 0) \leqslant \log 2-(1-\beta)\log(1-\beta) - \log(1+\beta) + \log(d-1+\delta)-\beta\log(\beta d-\beta+\delta), \quad \ldots (51)$$

where

$$\delta = +\sqrt{[\beta^2 d(d-2)+1]}. \quad \ldots (52)$$

Proof of Theorem 9 : The definition of $H_n^*(\beta, 0)$ shows that it is a non-increasing function of β; and then (26) implies the same for $\theta(\beta, 0)$. Because $\theta(\beta, 0)$ is non-increasing and concave for $0 \leqslant \beta \leqslant 1$, by virtue of Theorem 5, it must be continuous at $\beta = 0$. By (25), $\theta(\beta, 0)$ is non-negative. The continuity of $\theta(\beta, 0)$ at $\beta = 1$ follows from this and from (51), whose right-hand side tends to zero as β tends to 1 from below. Consequently, it only remains to prove (51), and this is merely the special case $n = 1$ of (45). For, when $n = 1$ all the Pólya walks are automatically self-avoiding. Hence

$$2d \exp\{K_1(t)\} = 2(d-1)+2\cosh t; \quad \ldots (53)$$

and an elementary calculation shows that

$$\mu_1(\beta) = K_1(\tau)-\beta\tau, \quad \ldots (54)$$

where τ is the root of

$$\sinh \tau = \beta(d-1+\cosh\tau). \quad \ldots (55)$$

On eliminating τ from (54) and (55) and substituting into (45), we get (51).

Theorem 10 :

$$\mu_n(n\beta)+\log f_n \leqslant n\theta(\beta, 0)+\gamma\sqrt{n}+\log[d(2n+1)]. \quad \ldots (56)$$

Proof of Theorem 10 : From (43) and the definitions of $K_n(t)$ and $H_n(t)$ we have

$$\begin{aligned} f_n \exp\{\mu_n(n\beta)\} &= f_n \exp\{\inf_t [K_n(t)-n\beta t]\} = \inf_t e^{-n\beta t} H_n(t) \\ &= \inf_t \sum_{x=-n}^{n} e^{(x-\beta)t} f_n(x) \leqslant \inf_{t \geqslant 0} \sum_{x=-n}^{n} e^{(x-n\beta)t} f_n(x) \\ &\leqslant (2n+1) \inf_{t\geqslant 0} \sup_{x \geqslant 0} e^{(x-n\beta)t} f_n(x). \quad \ldots (57) \end{aligned}$$

In the last step of (57) the factor $(2n+1)$ is the number of terms in the sum $\sum_{x=-n}^{n}$; and the supremum should accordingly be over $-n \leqslant x \leqslant n$. However, we may ignore the negative values of x, since $t \geqslant 0$ and $f_n(x)$ is an even function of x. From (57)

$$\begin{aligned} f_n \exp\{\mu_n(n\beta)\} &\leqslant (2n+1) \inf_{t \geqslant 0} \sup_{x \geqslant 0} e^{(x-n\beta)t} \sum_{y \geqslant x} f_n(y) \\ &= (2n+1) \inf_{t \geqslant 0} \sup_{x \geqslant 0} e^{(x-n\beta)t} H_n(x/n, 0). \quad \ldots (58) \end{aligned}$$

In (58) put $x = n\alpha$, where we may suppose $0 \leqslant \alpha \leqslant 1$. Thus

$$\begin{aligned} f_n \exp\{\mu_n(n\beta)\} &\leqslant (2n+1) \inf_{t \geqslant 0} \sup_{0 \leqslant \alpha \leqslant 1} e^{(\alpha-\beta)nt} H_n(\alpha, 0) \\ &\leqslant (2n+1)d \inf_{t \geqslant 0} \sup_{0 \leqslant \alpha \leqslant 1} e^{(\alpha-\beta)nt+n\theta(\alpha, 0)+\gamma\sqrt{n}} \quad \ldots (59) \end{aligned}$$

by virtue of (27). Consequently

$$\log f_n+\mu_n(n, \beta)-\gamma\sqrt{n}-\log[(2n+1)d] \leqslant n \inf_{t \geqslant 0} \sup_{0 \leqslant \alpha \leqslant 1} [(\alpha-\beta)t+\theta(\alpha, 0)] \leqslant n \sup_{0 \leqslant \alpha \leqslant 1} [\theta(\alpha, 0)+(\alpha-\beta)(-\theta'_\beta)], \quad \ldots (60)$$

where θ'_β is a one-sided derivative of $\theta(\beta, 0)$ at β. We may take, for definiteness, a right-hand derivative unless $\beta = 1$. The concavity of $\theta(\beta, 0)$ ensures the existence of θ'_β; and $-\theta'_\beta \geqslant 0$ since $\theta(\beta, 0)$ is non-increasing. Hence the last step in (60) merely consists in replacing t by a particular value, thus not decreasing the right-hand side. However, the concavity of $\theta(\beta, 0)$ implies that

$$\theta(\beta, 0)+(\alpha-\beta)\theta'_\beta \geqslant \theta(\alpha, 0). \quad \ldots (61)$$

Therefore

$$\log f_n+\mu_n(n\ \beta)-\gamma\sqrt{n}-\log[(2n+1)d] \leqslant n \sup_{0 \leqslant \alpha \leqslant 1} \theta(\beta, 0) = n\theta(\beta, 0), \quad \ldots (62)$$

which completes the proof of Theorem 10.

REFERENCES

BLACKWELL, D. and HODGES, J. L. (1959): The probability in the extreme tail of a convolution. *Ann. Math. Stat.*, **30**, 1113-1120.

CASASSA, E. F. (1960): Polymer solutions. *Ann. Rev. Phys. Chem.*, **11**, 477-500.

HAMMERSLEY, J. M. (1961): On the rate of convergence to the connective constant of the hypercubical lattice. *Oxford Q. J. Math.*, **12**(2), 250-256.

——— (1962): Generalization of the fundamental theorem on subadditive functions. *Proc. Camb. Phil. Soc.*, **58**, 235-238.

HAMMERSLEY, J. M. and WELSH, D. J. A. (1962): Further results on the rate of convergence to the connective constant of the hypercubical lattice. *Oxford Q. J. Math.*, **13**(2), 108-110.

HARDY, G. H., LITTLEWOOD, J. E., and PÓLYA, G. (1934): *Inequalities*, Cambridge University Press.

HARDY, G. H. and RAMANUJAN, S. (1917): Asymptotic formulae for the distribution of integers of various types. *Proc. London Math. Soc.*, **16**(2), 112-132.

HERMANS, J. J. (1957): High polymers in solution. *Ann. Rev. Phys. Chem.*, **8**, 179-198.

HILLE, E. (1948): *Functional Analysis and Semi-groups.* Amer. Math. Soc. Colloq. Publ., **31**.

MARCER, P. J. (1961): D. Phil. Thesis: Oxford University.

MARTIN, J. L. (1962): The exact enumeration of self-avoiding walks on a lattice. *Proc. Camb. Phil. Soc.*, **58**, 92-101.

O'FLAHERTY, M. P. (1961): Computations for the excluded volume problem in connection with the conformation of polymer molecules. *Stanford Univ. Appl. Math. and Stat. Lab. Tech. Rep.*, No. 11.

RENNIE, B. C. (1961): Random walks. *Magyar Tud. Akad. Mat. Kut. Intez. Közl.*, **6**, 263-269.

SYKES, M. E. (1961): Some counting theorems in the theory of the Ising model and the excluded volume problem. *J. Mathematical Phys.*, **2**, 52-62.

WALL, F. T. and HILLER, L. A. (1954): Properties of macromolecules in solution. *Ann. Rev. Phys. Chem.*, **5**, 267-290.

WALL, F. T., HILLER, L. A. and ATCHISON, W. F. (1955): Statistical computation of mean dimensions of macromolecules. *J. Chem Phys.*, **23**, 913-921, 2314-2321.

WALL, F. T. and ERPENBECK, J. J. (1959a): New method for the statistical computation of polymer dimensions. *J. Chem. Phys.*, **30**, 634-637.

——— (1959b): Statistical computation of radii of gyration and mean internal dimensions of polymer molecules. *J. Chem. Phys.*, **30**, 637-640.

Paper received : November, 1962.

A SIMPLE STOCHASTIC MODEL OF CONTINUOUS CULTURE OF MICROORGANISMS IN SEVERAL BASINS

By JAROSLAV KOŽEŠNIK

Academy of Sciences, Czechoslovakia

SUMMARY. The expected values and variances of the number of microorganism in cultivation basins and the formulae both for the steady as well as for the transient state are given. The condition for the culture to stabilise on the non-zero expected values has also been calculated.

1. Introduction

The purpose of this paper is to give a stochastic model of the development of a population of microorganisms when cultured in several subsequently joined basins through which the cultivation fluid flows. We assume that the external conditions (e.g. the temperature, the contents of nourishing substances and salts) are kept constant, in particular independent of time, for all the basins. However, one can also assume other conditions, in particular time-dependent conditions. In this paper we study the development of the population in question starting from a given initial number of microorganisms in the first basin till the stable (steady, stationary) case, taking into consideration conditions under which the stable state can be reached.

The basic idea is to study the development of the population as a multiple-parameter, complex branching process of birth, aging, migration and death. Under the notion birth we understand the division into two or four (multiple birth) new organisms. This happens with two different known probabilities, of course, only after the organism becomes mature.

The problem of maturing is expressed in such a way that we distinguish in the whole population, in all basins, between two groups of microorganisms, immature and mature ones, and the variation of the number of individuals in each of these two groups N_1, N_2 is studied. The number of such groups need not be limited by two. However, assuming a greater number of groups the computation of higher order moments becomes tedious. Introducing two groups, immature and mature, the process in question becomes an age-dependent branching stochastic process. It may be doubtful, whether the age-dependence in the process is expressed by means of distinguishing between two groups in a sufficient way. Some conclusions can be drawn when taking into consideration results obtained by Bellman, Harris and Bharucha-Reid (1960).

The equation for the first moment $\overline{N}(t)$ of the one-dimensional, age-dependent branching process reads :

$$\overline{N}(t) = 1-G(t)+\overline{K}\int_0^t \overline{N}(t-\tau)\,g(\tau)\,d\tau \qquad \ldots \quad (1.1)$$

where $G(t)$ is the distribution function of generation times, so that

$$g(t)dt = \frac{d}{dt} G(t)dt$$

is the probability that a single individual born in time t is subject to generation changes during the time interval $(t, t+dt)$. A generation change means the division of a single individual into n individuals with probabilities q^n, where $\sum_n q^n = 1$, so that

$$\overline{K} = \sum_n nq^n \quad \ldots \quad (1.2)$$

is the average number of new organisms into which the individual divides.

A well-known useful tool for solving (1.1) is the Laplace integral.

Thus
$$N(t) \doteqdot \frac{1-\hat{G}(p)}{1-\overline{K}\hat{G}(p)} \quad \ldots \quad (1.3)$$

where
$$\hat{G}(p) = p \int_0^{\infty} e^{-pt} G(t)dt.$$

By means of (1.1) or (1.3) we can study the variation of the expected value $\overline{N}(t)$ with varying distribution function $G(t)$ of generation times. The functions which seem to be very appropriate are of the type

$$G(t) = a^n \int_0^t \frac{\tau^{n-1}}{(n-1)!} e^{-a\tau} dt \quad \ldots \quad (1.4)$$

which lead to
$$\hat{G}(p) = \frac{a^n}{(p+a)^n}. \quad \ldots \quad (1.5)$$

The function $g(t)$ takes in this case its external value for $t = \frac{n-1}{a}$. Let us limit ourselves for a while to the case $n = 1$ (exponential distribution of generation times).

We have by (1.3)
$$\overline{N}(t) \doteqdot \frac{p}{p-a(\overline{K}-1)} \quad \ldots \quad (1.6)$$

and hence
$$\overline{N}(t) = e^{a(\overline{K}-1)t}. \quad \ldots \quad (1.7)$$

The exponent is positive, whenever $\overline{K} > 1$, and the population grows exponentially with time.

Now, let us compare this case with another one, when the generation change takes place always suddenly at a fixed age t_0 and cannot take place at other ages.

Then
$$\hat{G}(\rho) = e^{-\rho t_0} \qquad \dots \quad (1.8)$$

and therefore
$$\bar{N}(t) \doteqdot \frac{1-e^{-\rho t_0}}{1-\bar{K}e^{-\rho t_0}} = (1-e^{-\rho t_0})(1+\bar{K}e^{-\rho t_0}+\dots). \qquad \dots \quad (1.9)$$

The curve (1.9) is of "exponential" character, or, more precisely, oscillates along an exponential curve. Hence it follows, that even so trivial function $\hat{G}(\rho)$ as given by (1.8) has no essential influence on behaviour of $\bar{N}(t)$, provided we do not take into consideration short-term fluctuations.

For an arbitrary integer n we have by (1.3) and (1.5)

$$\text{(a)} \quad \bar{N}(t) \doteqdot \frac{(\rho+a)^n - a^n}{(p+a)^n - \bar{K}a^n}.$$

Let us put $\rho+a = q$

then we get
$$\text{(b)} \quad \nu(t) \doteqdot \frac{q^n - a^n}{q^n - \bar{K}a^n}$$

and also
$$\text{(c)} \quad \bar{N}(t) = \nu(t)e^{-at} + a\int_0^t \nu(\tau)\, e^{-a\tau}\, d\tau.$$

Determining the function $\nu(t)$ by (1.5) we determine at the same time, the function $\bar{N}(t)$. The function $\nu(t)$ can be determined e.g. by the partial fraction expansion method of (b). Clearly, we have
$$\text{(d)} \quad \nu(t) = \frac{1}{\bar{K}}\left(1 + \frac{\bar{K}-1}{n} \sum_{k=1}^{n} e^{q_k t} \right)$$

where
$$qk = a\sqrt[n]{\bar{K}}\left[\cos\frac{2\pi}{n}(k-1) + j\sin\frac{2\pi}{n}(k-1)\right].$$

Now, we can express the function $\bar{N}(t)$ for an arbitrary integer n. Since, with the only exception of $n = 1$ and $n = 2$, the roots of qk are always complex, $\bar{N}(t)$ is of damped-wave shape.

The probability density function of generation times reads in this case

$$g(t) = \frac{a^n t^{n-1}}{(n-1)!}\, e^{-at}.$$

Since the maximum of $g(t)$ is reached for $t = \frac{n-1}{a}$, we can expect that $\bar{N}(t)$ is of oscillation character mainly if the division takes place at old age only.

Now, let us give the formulae for the functions $g(t)$ and $\nu(t)$ for $n = 2, 3, 4$.

$$n = 2 \begin{cases} g(t) = a^2 t\, e^{-at} \\[2ex] \nu(t) = \dfrac{1}{\bar{K}} [1+(\bar{K}-1) \cos at \sqrt{\bar{K}}] \end{cases}$$

$$n = 3 \begin{cases} g(t) = \dfrac{a^3 t^2}{2!}\, e^{-at} \\[2ex] \nu(t) = \dfrac{1}{\bar{K}} \left[1+ \dfrac{K-1}{3} \left(e^{at\sqrt[3]{\bar{K}}} + 2e^{-at\sqrt[3]{\bar{K}}} \cos \dfrac{at\sqrt{3}\sqrt[3]{\bar{K}}}{2} \right)\right] \end{cases}$$

$$n = 4 \begin{cases} g(t) = \dfrac{a^4 t^3}{3\,!}\, e^{-at} \\[2ex] \nu(t) = \dfrac{1}{\bar{K}} \left[1+ \dfrac{\bar{K}-1}{2} \left(\cos at \sqrt[4]{\bar{K}} + \cos \text{at} \sqrt[4]{\bar{K}} \right)\right]. \end{cases}$$

The damped-wave character of oscillations for $n \geqslant 3$ is evident. The frequency of oscillations increases with increasing $\bar{K}a^n$. After $\nu(t)$ is determined, the function $\bar{N}(t)$ is given by (c).

According to the values of particular parameters in question, different cases can take place. Thus, e.g. for $n = 2$ the function $\bar{N}(t)$ decreases exponentially with increasing t whenever $a > \bar{K}$; and it increases otherwise.

In what follows we assume that after reaching the group 2 the probability of a generation change is the same for all individuals belonging to this group, be it the case of a single or a multiple birth. Thus, the process makes a branch. In general, different branches correspond to different probabilities.

2. Culture "fourterminal"

The culture fourterminal (see Fig. 1 on p. 63) is a basin filled with a fluid, in which some microorganisms live; these are according to some criterion, e.g. according to their age, divided into two groups. Let the number of individuals in each group be dependent on time t and equal to $N_1(t)$ and $N_2(t)$, respectively. Microorganisms migrate to the basin from the left with probability parameters equal to $N_1'(t)$ and $N_2'(t)$ assuming that the probability that a single individual will migrate from the left to the basin in the interval $(t, t+\Delta t)$ is equal to

$$N_1'(t)D'\Delta t+o(\Delta t) \quad \text{for type (1)}$$
$$N_2'(t)D'\Delta t+o(\Delta t) \quad \text{for type (2)}$$

where

$$\lim o(\Delta t) = 0 \text{ for } \Delta t \to 0.$$

The microorganisms migrate from the basin to the right in such a way, that the probability that a single individual will migrate from the basin to the right in the interval $(t, t+\Delta t)$ is equal to :

$$N_1(t)D''\Delta t+o(\Delta t) \quad \text{for type (1)}$$
$$N_2(t)D''\Delta t+o(\Delta t) \quad \text{for type (2).}$$

Organisms of both the groups can also die. The probability that a single individual will die in the interval $(t, t+\Delta t)$ is equal to

$$N_1(t)\delta_1\,\Delta t+o(\Delta t) \qquad \text{for type (1)}$$
$$N_2(t)\delta_2\Delta t+o(\Delta t) \qquad \text{for type (2).}$$

Organisms of the first group mature during their stay in the basin and thus pass to the second group. The probability, that a single individual passes from the first group to the second group in the interval $(t, t+\Delta t)$ is equal to

$$N_1(t)\mu_{1,2}\Delta t+o(\Delta t).$$

The generation change takes place only in the second group. The probability that a single individual of the second group divides into two new individuals of the first group in the interval $(t, t+\Delta t)$ is equal to

$$N_2(t)\mu_{2,2}\Delta t+o(\Delta t)$$

and in a similar way the probability that a multiple birth takes place in the interval $(t, t+\Delta t)$ is equal to

$$N_2(t)\mu_{2,4}\Delta t+o(\Delta t).$$

The particular probability parameters for entrance to and exit from the basin D', D'' can be influenced mechanically, i.e. by a different choice of the diameter and the placing of the pipes under consideration, and also by changing the speed of the fluid. In the first approximation we can write

$$D=\frac{fv}{V}$$

where v is the velocity in the connection pipes, f the area of the orifice, and V the volume of the basin.

On the contrary δ_1, δ_2, $\mu_{2,2}$, $\mu_{2,4}$, $\mu_{1,2}$ can be influenced mainly by a proper choice of external conditions (the temperature, the contents of nourishing substances in the fluid, the amount of the absorbed light, etc.). Now, our task is to write down the fundamental equation for $P(N_1', N_2', N_1, N_2, t)$, i.e. for the marginal probability that in the moment t the number of organisms of both the types in the basin in question is equal to $N_1(t)$ and $N_2(t)$ while the number of organisms of both the types in the preceding basins is equal to $N_1'(t)$ and $N_2'(t)$, respectively. Having started from the above mentioned elementary probabilities and proceeding in a usual way, we get the equation

$$\begin{aligned}
&\frac{\partial}{\partial t}\,P(N_1', N_2', N_1, N_2, t)\\
&=[P(N_1', N_2', N_1+1, N_2, t)(N_1+1)-P(N_1', N_2', N_1, N_2, t)N_1]\delta_1\\
&\quad+[P(N_1', N_2', N_1, N_2+1, t)(N_2+1)-P(N_1', N_2', N_1, N_2, t)N_2]\delta_2\\
&\quad+[P(N_1'+1, N_2', N_1-1, N_2, t)(N_1'+1)-P(N_1', N_2', N_1, N_2, t)N_1']D'\\
&\quad+[P(N_1', N_2'+1, N_1, N_2-1, t)(N_2'+1)-P(N_1', N_2', N_1, N_2, t)N_2']D'\\
&\quad+[P(N_1', N_2', N_1+1, N_2, t)(N_1+1)-P(N_1', N_2', N_1, N_2, t)N_1]D''\\
&\quad+[P(N_1', N_2', N_1, N_2+1, t)(N_2+1)-P(N_1', N_2', N_1, N_2, t)N_2]D''\\
&\quad+[P(N_1', N_2', N_1-2, N_2+1, t)(N_2+1)-P(N_1', N_2', N_1, N_2, t)N_2]\mu_{2,2}\\
&\quad+[P(N_1', N_2', N_1-4, N_2+1, t)(N_2+1)-P(N_1', N_2', N_1, N_2, t)N_2]\mu_{2,4}\\
&\quad+[P(N_1', N_2', N_1+1, N_2-1, t)(N_1+1)-P(N_1', N_2', N_1, N_2, t)N_1]\mu_{1,2}] \qquad \dots \quad (2.1)
\end{aligned}$$

From this equation we determine first and second order moments of particular random variables, i.e. $\bar{N}_1(t)$ and $\bar{N}_2(t)$ or $\bar{\bar{N}}_1(t)$ and $\bar{\bar{N}}_2(t)$, and all covariances of interest, that characterise the population in the basin. To find an explicit solution of equation (2.1) is not easy. However, the moments can be easily determined (see Feller, 1958, p. 411). For first moments we get equations

$$\frac{\partial}{\partial t}\bar{N}_1(t) = -\bar{N}_1(t)(\delta_1+D''+\mu_{1,2})+2\bar{N}_2(t)(\mu_{2,2}+2\mu_{2,4})+\bar{N}_1'(t)D' \qquad \dots \quad (2.2a)$$

and $$\frac{\partial}{\partial t}\bar{N}_2(t) = \bar{N}_1(t)\mu_{1,2}-\bar{N}_2(t)(\delta_2+D''+\mu_{2,2}+\mu_{2,4})+\bar{N}_2'(t)\,D' \qquad \dots \quad (2.2b)$$

We apply Laplace-Wagner integral to both the equations, setting $\bar{N}_1(t) = \bar{N}_{1,0}$ and $\bar{N}_2(t) = N_{2,0}$ for $t = 0$. We get

$$\begin{Vmatrix} \hat{N}_1'(\rho) \\ \hat{N}_2'(\rho) \end{Vmatrix} = \frac{1}{D'}\begin{Vmatrix} \rho+\delta_1+D''+\mu_{1,2}, & -2(\mu_{2,2}+2\mu_{2,4}) \\ -\mu_{1,2}, & \rho+\delta_2+D''+\mu_{1,2}+\mu_{2,4} \end{Vmatrix}\begin{Vmatrix} \hat{N}_1(\rho) \\ \hat{N}_2(\rho) \end{Vmatrix} - \frac{\rho}{D'}\begin{Vmatrix} \bar{N}_{1,0} \\ \bar{N}_{2,0} \end{Vmatrix} \dots \quad (2.3)$$

which is the fundamental equation for calculation of expected values $\bar{N}_1(t)$ and $\bar{N}_2(t)$ of the cultivation fourterminal. The Laplace operator was denoted in equation (2.3) by ρ.

From equation (2.3) we can derive some interesting results for one-side-isolated cultivation fourterminal, which is defined by $N_1'(t) = 0$ and $N_2'(t) = 0$ for all t. Then we have

$$\begin{Vmatrix} \hat{N}_1(\rho) \\ \hat{N}_2(\rho) \end{Vmatrix} = \frac{\rho}{\Delta}\begin{Vmatrix} \rho+\delta_2+D''+\mu_{2,2}+\mu_{2,4}, & 2(\mu_{2,2}+2\mu_{2,4}) \\ \mu_{1,2}, & \rho+\delta_1+D''+\mu_{1,2} \end{Vmatrix}\begin{Vmatrix} \bar{N}_{1,0} \\ \bar{N}_{2,0} \end{Vmatrix} \qquad \dots \quad (2.4)$$

where Δ is the determinant of the square matrix occurring in (2.3). Evidently, the influence of D' disappears. Determinant Δ is a quadratic function in ρ. After multiplying we get

$$\Delta = (\rho+\delta_1+D''+\mu_{1,2})(\rho+\delta_2+D''+\mu_{2,2}+\mu_{2,4})-2\mu_{1,2}(\mu_{2,2}+2\mu_{2,4}).$$

Let us ask when the limits of (2.4) for $\rho\to 0$ exist and do not equal 0 (the limit for $\rho\to 0$ corresponds to the limit for $t\to\infty$). (Tauber's Theorem; Pol-Bremmer (1950)),

Evidently, this case takes place if

$$(\delta_1+D''+\mu_{1,2})(\delta_2+D''+\mu_{2,2}+\mu_{2,4})-2\mu_{1,2}(\mu_{2,2}+\mu_{2,4}) = 0. \qquad \dots \quad (2.5)$$

Namely, we have then

$$\Delta = \rho(\rho+\delta_1+2D''+\mu_{1,2}+\mu_{2,2}+\mu_{2,4})$$

so that we can reduce the powers of ρ in (2.4) and then we can set $\rho = 0$ and then determine the limit values $\bar{N}_1(t)$ and $\bar{N}_2(t)$.

Hence, equation (2.5) represents the condition that in an isolated culture fourterminal the non-zero stable state can be reached. From the expression (2.3) for first order moments of a general culture fourterminal, we can easily derive expres-

sions for special fourterminals. Thus, for example, for the reception twoterminal, i.e. for the reservoir defined by

$$\delta_1 = 0,\ D'' = 0,\ \mu_{1,2} = 0,\ \mu_{2,2} = 0,\ \mu_{2,4} = 0,\quad \delta_2 = 0,\ \bar{N}_{1,0} = 0,\ \bar{N}_{2,0} = 0,$$

we get
$$\hat{N}_1(\rho) = \frac{D'}{\rho}\hat{N}_1'(\rho);\quad \hat{N}_2(\rho) = \frac{D'}{\rho}\hat{N}_2'(\rho)$$

so that we have
$$\bar{N}_1(t) = D'\int_0^t \bar{N}_1'(\tau)d\tau;\quad \bar{N}_2(t) = D'\int_0^t \bar{N}_2'(\tau)d\tau. \qquad \ldots \quad (2.6)$$

In a similar way we can proceed even in other cases when the particular probability parameters have special values. Also, for second order moments of the fourterminal the necessary expressions can be derived. However, the derivation is more tedious, because not only second order moments of $N_1'(t)$, $N_2'(t)$, $N_1(t)$, $N_2(t)$, but also corresponding covariances $\mathrm{cov}(N_1, N_2)$, $\mathrm{cov}(N_1, N_1')$, $\mathrm{cov}(N_2, N_1')$, $\mathrm{cov}(N_2', N_1)$, $\mathrm{cov}(N_2', N_2)$, must be used for calculation.

Altogether we have to deal with mutual relations of seven quantities.

Under the same assumption as for first order moments we can derive from equation (2.1) the following simultaneous differential equations for second order moments.

(a)
$$\frac{\partial}{\partial t}\bar{\bar{N}}_1 = -2\bar{\bar{N}}_1(\delta_1+D''+\mu_{1,2})+4(\mu_{2,2}+2\mu_{2,4})\,\mathrm{cov}(N_1, N_2)+2D'\,\mathrm{cov}(N_1, N_1')$$
$$+\bar{N}_1(\delta_1+D''+\mu_{1,2})+4\bar{N}_2(\mu_{2,2}+4\mu_{2,4})+D'\bar{N}_1'$$

(b)
$$\frac{\partial}{\partial t}\bar{\bar{N}}_2 = -2\bar{\bar{N}}_2(\delta_2+D''+\mu_{2,2}+\mu_{2,4})+2\mu_{1,2}\,\mathrm{cov}(N_1, N_2)+2D'\,\mathrm{cov}(N_2, N_2')$$
$$+\bar{N}_2(\delta_2+D''+\mu_{2,2}+\mu_{2,4})+\bar{N}_1\mu_{1,2}+D'\bar{N}_2'$$

(c)
$$\frac{\partial}{\partial t}\mathrm{cov}(N_1, N_2) = -\,\mathrm{cov}(N_1, N_2)(\delta_1+\delta_2+2D''+\mu_{1.2}+\mu_{2,2}+\mu_{2,4})+D'\,\mathrm{cov}(N_1, N_2')$$
$$+D'\,\mathrm{cov}(N_1', N_2)+\mu_{1,2}\bar{\bar{N}}_1+2\bar{\bar{N}}_2(\mu_{2,2}+2\mu_{2,4})$$
$$-\bar{N}_1\mu_{1,2}-2\bar{N}_2(\mu_{2,2}+2\mu_{2,4})$$

(d)
$$\frac{\partial}{\partial t}\mathrm{cov}(N_1', N_1) = -\mathrm{cov}(N_1', N_1)(\delta_1+D'+D''+\mu_{1,2})$$
$$+2(\mu_{2,2}+2\mu_{2,4})\,\mathrm{cov}(N_1', N_2)+D'\bar{\bar{N}}_1'-D'\bar{N}_1'$$

(e)
$$\frac{\partial}{\partial t}\mathrm{cov}(N_1', N_2) = -\,\mathrm{cov}(N_1', N_2)(\delta_2+D'+D''+\mu_{2,2}+\mu_{2,4})$$
$$+D'\,\mathrm{cov}(N_1', N_2')+\mu_{1,2}\,\mathrm{cov}(N_1, N_1')$$

(f) $$\frac{\partial}{\partial t}\operatorname{cov}(N_1, N_2') = -\operatorname{cov}(N_1, N_2')(\delta_1+D'+D''+\mu_{1,2}) + D'\operatorname{cov}(N_1', N_2') + 2\operatorname{cov}(N_2, N_2')(\mu_{2,2}+2\mu_{2,4})$$

(g) $$\frac{\partial}{\partial t}\operatorname{cov}(N_2, N_2') = -\operatorname{cov}(N_2, N_2')(\delta_2+D'+D''+\mu_{2,2}+\mu_{2,4}) + \mu_{1,2}\operatorname{cov}(N_1, N_2') + D'\bar{\bar{N}}_2' - D'\bar{N}_2'. \qquad \ldots \quad (2.7)$$

These relatively complicated equations are again transformed by means of Laplace integral. Without any loss of generality we can suppose that for $t = 0$ all second order moments equal 0. In such a way further calculations simplify a little and we can immediately write a matrix relation, which resembles relation (2.3), namely

$$\left\| \begin{array}{c} \bar{\bar{N}}_1 \\ \bar{\bar{N}}_2 \\ \operatorname{cov}(N_1, N_2) \\ \operatorname{cov}(N_1', N_1) \\ \operatorname{cov}(N_1', N_2) \\ \operatorname{cov}(N_2', N_1) \\ \operatorname{cov}(N_2', N_2) \end{array} \right\| = B^{-1}(\rho)L \left\| \begin{array}{c} \bar{N}_1' \\ \bar{N}_2 \\ \operatorname{cov}(N_1', N_2') \\ \bar{\bar{N}}_1' \\ \bar{\bar{N}}_2' \\ \bar{N}_1 \\ \bar{N}_2 \end{array} \right\| \qquad \ldots \quad (2.8)$$

where both $B(p)$ and L are square matrices 7 by 7. We have $B(\rho) =$

$$\left| \begin{array}{ccccccc} \rho+2(\delta_1+D''+\mu_{1,2}) & 0 & -4(\mu_{2,2}+2\mu_{2,4}) & -2D' & 0 & 0 & 0 \\ 0 & \rho+2(\delta_2+D''+\mu_{2,2}+\mu_{2,4}) & -2\mu_{1,2} & 0 & 0 & 0 & -2D' \\ -\mu_{1,2} & -2(\mu_{2,2}+2\mu_{2,4}) & \rho+\delta_1+\delta_2+2D''+\mu_{1,2}+\mu_{2,2}+\mu_{2,4} & 0 & -D' & -D' & 0 \\ 0 & 0 & 0 & \rho+\delta_1+D'+D''+\mu_{1,2} & -2(\mu_{2,2}+2\mu_{2,4}) & 0 & 0 \\ 0 & 0 & 0 & -\mu_{1,2} & \rho+\delta_2+D'+D''+\mu_{2,2}+\mu_{2,4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho+\delta_1+D'+D''+\mu_{1,2} & -2(\mu_{2,2}+2\mu_{2,4}) \\ 0 & 0 & 0 & 0 & 0 & -\mu_{1,2} & \rho+\delta_2+D'+D''+\mu_{2,2}+\mu_{2,4} \end{array} \right|$$

$$L = \begin{vmatrix} D' & 0 & 0 & 0 & 0 & \delta_1+D''+\mu_{1,2} & 4(\mu_{2,2}+4\mu_{2,4}) \\ 0 & D' & 0 & 0 & 0 & \mu_{1,2} & \delta_2+D''+\mu_{2,2}+\mu_{2,4} \\ 0 & 0 & 0 & 0 & 0 & -\mu_{1,2} & -2(\mu_{2,2}+2\mu_{2,4}) \\ -D' & 0 & D' & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & D' & 0 & 0 \\ 0 & 0 & 0 & 0 & D' & 0 & 0 \\ 0 & -D' & 0 & D' & 0 & 0 & 0 \end{vmatrix}$$

The matrix L consists mostly of zeros.

Equation (2.8) enables us to compute all second order moments of the population in the basin, provided we know both moments of the immigrating population and first order moments of the population in the basin. The calculations are cumbersome, however, possible. The stable state can be considered in a similar way as it was when dealing with first order moments case. The determinant of the matrix $B(\rho)$ is a polynomial in ρ of the seventh degree. On the contrary, L does not depend on ρ. The limits of second order moments for $t\to\infty$, provided such limits exist, can be derived when limiting (2.8) for $\rho\to 0$. The computation of such limit values is relatively easy.

3. Closed chain of culture fourterminals

For the sake of continuity we join single culture fourterminals into a continual closed chain. This section is therefore devoted to the derivation of expected values $\bar{N}_1(t)$ and $\bar{N}_2(t)$ in particular links of the chain. These values are in an unstable state dependent on time. In an equilibrium state they are stable. In what follows we confine ourselves to the first order moments, because the calculation of second order moments is essentially similar and, of course, very tedious.

Let us assume a chain as sketched in Fig. 2 on p. 63. It consists of a set of basins through which the fluid flows and carries the cultured microorganisms. Let us number the basins subsequently by numbers 1, 2, ..., . We shall also distinguish the values N_1 and N_2 by corresponding indices in such a way that $N_{1,1}$ and $N_{2,1}$ are the instantaneous numbers of microorganisms in the first and the second group of the first basin, respectively. In a similar way we proceed in other basins, in general we write $N_{1,i}$ and $N_{2,i}$. In an analogous way we distinguish between corresponding moments, say for example $\bar{N}_{1,i}$ and $\bar{N}_{2,i}$. Let us assume that at the time instant $t = 0$ the fluid in all the basins contains no microorganisms. At the same time instant $N_{1,0}$ microorganisms of the first group and $N_{2,0}$ microorganisms of the second group are supplied into the first basin only and the process of cultivation starts. If the conditions for reaching the stable state are fulfilled, then this state is really reached after some time

(theoretically after an infinite time). By (2.3) we get for the i-th basin the following abbreviated expression

$$\left\| \begin{matrix} \hat{N}'_{1,i}(\rho) \\ \hat{N}'_{2,i}(\rho) \end{matrix} \right\| = \frac{a_i}{D'_i} \left\| \begin{matrix} \hat{N}_{1,i}(\rho) \\ \hat{N}'_{2,i}(\rho) \end{matrix} \right\| - \frac{\rho}{D'_i} \left\| \begin{matrix} \overline{N}_{1,i}(0) \\ \overline{N}_{2,i}(0) \end{matrix} \right\| \qquad \ldots \quad (3.1)$$

Clearly we have also

$$\hat{N}'_{1,i+1}(\rho) = \hat{N}_{1,i}(\rho); \quad \overline{N}'_{1,i+1}(t) = \overline{N}_{1,i}(t) \qquad \ldots \quad (3.1a)$$

$$\hat{N}'_{2,i+1}(\rho) = \hat{N}_{2,i}(\rho); \quad \overline{N}'_{2,i+1}(t) = \overline{N}_{2,i}(t) \qquad \ldots \quad (3.1b)$$

$$D''_i = D'_{i+1}. \qquad \ldots \quad (3.1c)$$

According to the assumptions stated above, we have $\overline{N}_{1,i}(0) = 0$, $\overline{N}_{2,i}(0) = 0$ for all i except $i = 1$. Thus we get (omitting the operator ρ)

$$\left\| \begin{matrix} \hat{N}'_{1,1} \\ \hat{N}'_{2,1} \end{matrix} \right\| = \frac{a_1}{D'_1} \left\| \begin{matrix} \hat{N}_{1,1} \\ \hat{N}_{2,1} \end{matrix} \right\| - \frac{\rho}{D'_1} \left\| \begin{matrix} \overline{N}_{1,0} \\ \overline{N}_{2,0} \end{matrix} \right\|$$

$$\left\| \begin{matrix} \hat{N}_{1,1} \\ \hat{N}_{2,1} \end{matrix} \right\| = \frac{a_2}{D'_2} \left\| \begin{matrix} \hat{N}_{1,2} \\ \hat{N}_{2,2} \end{matrix} \right\|$$

and finally

$$\left\| \begin{matrix} \hat{N}_{1'n-1} \\ \hat{N}_{2,n-1}' \end{matrix} \right\| = \frac{a_n}{D'_n} \left\| \begin{matrix} \hat{N}_{1,n} \\ \hat{N}_{2,n} \end{matrix} \right\|$$

Eliminating subsequently all $\hat{N}_{1,i}$, $\hat{N}_{2,i}$ we get

$$\left\| \begin{matrix} \hat{N}'_{1,1} \\ \hat{N}'_{2,1} \end{matrix} \right\| = \frac{a_1 a_2 \ldots a_n}{D'_1 D'_2 \ldots D'_n} \left\| \begin{matrix} \hat{N}_{1,n} \\ \hat{N}_{2,n} \end{matrix} \right\| - \frac{\rho}{D'_1} \left\| \begin{matrix} \overline{N}_{1,0} \\ \overline{N}_{2,0} \end{matrix} \right\| \qquad \ldots \quad (3.2)$$

Since by the assumption $\hat{N}_{1,n} = N'_{1,1}$ and also $\hat{N}_{2,n} = \hat{N}'_{2,1}$ we can finally write

$$\left[\frac{a_1 a_2 \ldots a_n}{D'_1 D'_2 \ldots D'_n} - j \right] \left\| \begin{matrix} \hat{N}_{1,n} \\ \hat{N}_{2,n} \end{matrix} \right\| = \frac{\rho}{D'_1} \left\| \begin{matrix} \overline{N}_{1,0} \\ \overline{N}_{2,0} \end{matrix} \right\| \qquad \ldots \quad (3.3)$$

where j is the unit matrix of the second order. A special attention should be directed to the matrix a_n. Namely, the n-th fourterminal has two kinds of immigration (see Fig. 3 on p. 63) : either to the reservoir or to the first fourterminal (the feed-back).

If we denote by $D''_{a,n}$ the probability parameter of immigration to the reservoir and by $D''_{b,n}$ the probability parameter of immigration to the first fourterminal, so that

$$D''_n = D''_{a,n} + D''_{b,n} \qquad \ldots \quad (3.4)$$

then the matrix a_n of the n-th fourterminal will contain the whole probability parameter D''_n. By (2.6) growth of the number of microorganisms in the reservoir is then expressed by

$$\overline{N}_{1,R}(t) + \overline{N}_{2,R}(t) = D''_{a,n} \int_0^t (N_{1,n}(\tau) + N_{2,n}(\tau)) d\tau. \qquad \ldots \quad (3.5)$$

Evidently, we have then

$$D''_{b,n} = D'_1. \qquad \ldots \quad (3.6)$$

If the conditions in all the subsequent basins are the same, then also all the matrices a_i equal each other and (3.3) simplifies to

$$\left[\frac{a^n}{D'_1 D'_2 \ldots D_n} - j\right] \left\| \begin{matrix} \hat{N}_{1,n} \\ \hat{N}_{2,n} \end{matrix} \right\| = \frac{\rho}{D'_1} \left\| \begin{matrix} \bar{N}_{1,0} \\ \bar{N}_{2,0} \end{matrix} \right\| \qquad \ldots \quad (3.7)$$

The equation (3.3) enables us to compute the operator transforms $\hat{N}_{1,n}(\rho)$ and $\hat{N}_{2,n}(\rho)$ of the expected values $\bar{N}_{1,n}(t)$ and $\bar{N}_{2,n}(t)$, respectively.

Thus, being able to compute these values we can also compute from equation (3.5) the growth of the number of microorganisms in the reservoir. As to the first order moments, the whole problem of continual cultivation is thus solved. In a similar way we can proceed when computing second order moments using equation (2.8). However, computations are cumbersome and will be omitted in this paper.

On the contrary, we will consider the problem of determining such conditions for the closed chain of culture fourterminals that the process not only can reach the stable state but also stay in it.

4. Conditions for reaching the stable state in the culture system

These conditions follow from equation (2.2) in a similar way as in the case of one basin (see (2.2) and (2.3)). For the expected values $\bar{N}_{1,n}(t)$ and $\bar{N}_{2,n}(t)$ to stabilize on some non-zero values it is necessary and sufficient that the determinant of the final square matrix of the second order, i.e., of the matrix

$$\left[\frac{a_1 a_2 \ldots a_n}{D'_1 D'_2 \ldots D_n} - j\right] \qquad \ldots \quad (4.1)$$

has the value zero for $\rho \to 0$. Then it is possible, similarly as in the case of a single basin (see (2.4)), to reduce the power of the operator ρ, which is the condition for $\bar{N}_1(t)$ and $\bar{N}_2(t)$ to stabilize on non-zero values.

The general computation procedure is evident from equations (3.1), (3.2) and (3.3). Now, the question is to derive the stability condition and to determine the stable values $\lim_{t\to\infty} \bar{N}_1(t)$ and $\lim_{t\to\infty} \bar{N}_2(t)$ in all basins forming the closed chain and finally to determine regularity conditions that govern the culture during the transient period before the stable state is reached.

For determining average values at the end of the chain, i.e., the values $\bar{N}_{1,n}(t)$ and $\bar{N}_{2,n}(t)$, we have at our disposal equation (3.3) which enables us to compute corresponding Laplace transforms $\hat{N}_{1,n}(\rho)$ and $\hat{N}_{2,n}(\rho)$.

The first step is to compute the matrix product $a_1 a_2 \ldots a_n$, which contains the operator ρ. In order to get simple results let us introduce the following notation :

$$a_i = \begin{Vmatrix} \bar{a}_i+\rho, & \bar{b}_i \\ \bar{c}_i \;, & \bar{d}_i+\rho \end{Vmatrix} \qquad \ldots \quad (4.2)$$

where a_i, b_i, c_i and d_i are constants dependent on conditions in the i-th basin (cf.(2.2)). Let us calculate the eigenvalues $(\lambda_i)_{1,2}$ of the matrix a_i so that we set

$$\begin{vmatrix} \bar{a}_i-(\lambda_i-\rho), & \bar{b}_i \\ \bar{c}_i \quad , & \bar{d}_i-(\lambda_i-\rho) \end{vmatrix} = 0.$$

Hence
$$(\lambda_i-\rho)^2-(\lambda_i-\rho)(\bar{a}_i+\bar{d}_i)+\bar{a}_i\bar{d}_i-\bar{b}_i\bar{c}_i = 0 \qquad \ldots \quad (4.3)$$

so that
$$(\lambda_i)_{1,2} = \rho+\tfrac{1}{2}(\bar{a}_i+\bar{d}_i)\pm\sqrt{\tfrac{1}{4}(\bar{a}_i+\bar{d}_i)^2-(\bar{a}_i\bar{d}_i-\bar{b}_i\bar{c}_i)}. \qquad \ldots \quad (4.4)$$

Evidently the eigenvalues of the matrix a_i are linear functions of the operator ρ.

Let us now compute the product $v(\rho)$ of the matrices

$$v(\rho) = a_1\, a_2 \ldots a_n \qquad \ldots \quad (4.5)$$

where we get by (4.2)

$$a_i = \begin{Vmatrix} \bar{a}_i, & \bar{b}_i \\ \bar{c}_i, & \bar{d}_i \end{Vmatrix} + \rho \begin{Vmatrix} 1, & 0 \\ 0, & 1 \end{Vmatrix} = a_{0,i}+\rho j. \qquad \ldots \quad (4.5a)$$

Then we get

$$v(\rho) = (a_{0,1}+\rho j)(a_{0,2}+\rho j) \ldots (a_{0,n}+\rho j)$$

$$= a_{0,1}\, a_{0,2} \ldots a_{0,n}+\rho \sum_k \frac{a_{0,1}\, a_{0,2} \ldots a_{0,n}}{a_{0,k}}$$

$$+\rho^2 \sum_j \sum_k \frac{a_{0,1}\, a_{0,2} \ldots a_{0,n}}{a_{0,j}\, a_{0,k}} +\ldots+\rho^n j. \qquad \ldots \quad (4.6)$$

Let us remark that the expressions as $\frac{a_{0,1}\, a_{0,2} \ldots a_{0,n}}{a_{0,j}}$ or $\frac{a_{0,1}\, a_{0,2} \ldots a_{0,n}}{a_{0,j}\, a_{0,k}}$ are symbols only and by no means denote division by matrices $a_{0,j}$ or by the matrix product $a_{0,j}\ a_{0,k}$. These symbols are introduced for the sake of simplicity of writing only.

Let us rewrite (4.6) in the form

$$v(\rho) = \kappa_{0,0}+\rho\kappa_{0\,1}+\rho^2\kappa_{0,2}+\ldots+\rho^n j \qquad \ldots \quad (4.7)$$

where the symbols $\kappa_{0,i}$ are clear from comparison with (4.6).

The determinant of the matrix $v(\rho)$ is for $\rho \to 0$ equal to

$$|v(0) = |a_{0,1}\, a_{0,2} \ldots a_{0,n}| = |\kappa_{0,0}| \qquad \ldots \quad (4.8)$$

and can be computed as the product of determinants of matrices in question. Let us introduce the notation

$$\kappa_{0,i} = \begin{Vmatrix} x_{1,i}, & x_{2,i} \\ x_{3,i}, & x_{4,i} \end{Vmatrix} \qquad \ldots \quad (4.9)$$

Then we can write instead of (4.7) or (4.5) the matrix

$$v(\rho) = \left\| \begin{matrix} x_{1,0}+\rho x_{1,1}+\rho^2 x_{1,2}+\ldots, & x_{2,0}+\rho x_{2,1}+\rho^2 x_{2,2}+\cdots \\ x_{3,0}+\rho x_{3,1}+\rho^2 x_{\ ,2}+\ldots, & x_{4,0}+\rho x_{4,1}+\rho^2 x_{4,2}+\cdots \end{matrix} \right\| \qquad \ldots \quad (4.10)$$

Now let us determine the determinant of the matrix (4.1). Clearly we have

$$\left| \frac{a_1 a_2 \ldots a_n}{D_1' D_2' \ldots D_n'} - j \right| = \frac{1}{D_1' D_2' \ldots D_n'} \left| a_1 a_2 \ldots a_n - D_1' D_2 \ldots D_n j \right|.$$

Hence the determinant in question equals :

$$\Delta(\rho) = |v(\rho) - D_1' D_2' \ldots D_n' j| = \left| \begin{matrix} x_{1,0}+\rho x_{1,1}+\ldots - D_1' D_2' \ldots D_n', & x_{2,0}+\rho x_{2,1}+\cdots \\ x_{3,0}+\rho x_{3,1}+\ldots, & x_{4,0}+\rho x_{4,1}+\ldots - D_1' D_2' \ldots D_n' \end{matrix} \right| \qquad \ldots \quad (4.11)$$

With respect to the further need we will compute the coefficients at the zero-th and the first power of ρ only. We have

$$\begin{aligned} \Delta(\rho) = {} & (x_{1,0} - D_1' D_2' \ldots D_n')(x_{4,0} - D_1' D_2' \ldots D_n') - x_{2,0}\, x_{3,0} \\ & + \rho[x_{4,1}(x_{1,0} - D_1' D_2' \ldots D_n') + x_{1,1}(x_{4,0} - D_1' D_2' \ldots D_n') \\ & - x_{2,1}\, x_{3,0} - x_{3,1}\, x_{2,0}] + \cdots \end{aligned} \qquad \ldots \quad (4.12)$$

Evidently for $\rho = 0$ we have

$$\begin{aligned} \Delta(0) & = (x_{1,0} - D_1' D_2' \ldots D_n')(x_{4,0} - D_1' D_2' \ldots D_n') - x_{2,0} x_{3,0} \\ & = (D_1' D_2' \ldots D_n')^2 - D_1' D_2' \ldots D_n'(x_{1,0} + x_{4,0}) + x_{1,0}\, x_{4,0} - x_{2,0}\, x_{3,0} \\ & = (D_1' D_2' \ldots D_n')^2 - D_1' D_2' \ldots D_n' \text{ tr } v(0) + |v(0)|. \end{aligned} \qquad \ldots \quad (4.13)$$

In the expression (4.13) tr $v(0)$ denotes the trace of the matrix $v(0)$, i.e. $x_{1,0}+x_{4,0}$, and $|v(0)|$ the determinant of this matrix according to (4.8). Let us further denote the coefficient at the first power of ρ in the equation (4.12) by

$$\begin{aligned} P(0) = {} & x_{4,1}(x_{1,0} - D_1' D_2' \ldots D_n') + x_{1,1}(x_{4,0} - D_1' D_2' \ldots D_n') \\ & - x_{2,1}\, x_{3,0} - x_{3,1}\, x_{2,0} \end{aligned} \qquad \ldots \quad (4.14)$$

The condition of the stable state is that the determinant (4.13) equals zero, thus

$$(D_1' D_2' \ldots D_n')^2 - D_1' D_2' \ldots D_n' \text{ tr } v(0) + |v(0)| = 0. \qquad \ldots \quad (4.15)$$

This is necessary and sufficient for the culture process to stabilize. This condition simplifies when all the matrices a_i in the product (4.5) are equal, i.e. when conditions in all n basins are the same. The product (4.5) then becomes the n-th power of the matrix a and the stability condition simplifies by introducing the eigenvalues λ_1, λ_2 of this matrix (see (4.2) and (4.4), where the index should be deleted).

The eigenvalues of the matrix a^n are λ_1^n, λ_2^n, the determinant $|a^n| = \lambda_1^n \lambda_2^n$ and the trace $\text{tr}\, a^n = \lambda_1^n + \lambda_2^n$. Therefore (4.15) becomes

$$(D_1' D_2' \ldots D_n')^2 - D_1' D_2' \ldots D_n' (\lambda_1^n + \lambda_2^n) + \lambda_1^n \lambda_2^n = 0$$

which we can write in the form

$$(D_1' D_2' \ldots D_n' - \lambda_1^n)(D_1' D_2' \ldots D_n' - \lambda_2^n) = 0$$

so that finally either

$$D_1' D_2' \dots D_n' = \lambda_1^n \qquad \dots \quad (4.16a)$$

or

$$D_1' D_2' \dots D_n' = \lambda_2^n. \qquad \dots \quad (4.16b)$$

Thus, in the case of equal basins the stability condition is especially simple. Conditions (4.15) or (4.16a) and (4.16b) can be in the easiest way fulfilled so that we choose D_1' in a proper way. Let us remark that other parameters $D_2', D_3', \dots, D_n'$ are not free, they are dependent on $D_1'', D_2'', \dots$ (see (3.1c)) and they participate in matrices a_i. Further, for the solution to be practically realizable we must have

$$0 < D_1' < D_n''. \qquad \dots \quad (4.16c)$$

Therefore we always choose from two possible values, the value that ensures (4.16c).

5. Computation of stable states

We shall start from equation (3.3). Employing (4.5) we can write instead of (3.3)

$$[v(\rho) - D_1' D_2' \dots D_n' j] \begin{Vmatrix} \hat{N}_{1,n} \\ \hat{N}_{2,n} \end{Vmatrix} = D_2' D_3' \dots D_n' \rho \begin{Vmatrix} \overline{N}_{1,0} \\ \overline{N}_{2,0} \end{Vmatrix} \qquad \dots \quad (5.1)$$

To calculate $\hat{N}_{1,n}$ and $\hat{N}_{2,n}$ we need the determinant of the final matrix given by the sum of matrices in the braces on the left-hand side of this equation. However, according to (4.11) that equals determinant Δ. Let us denote $z = D_1' D_2' \dots D_n'$ and also

$$\Delta = \begin{vmatrix} \mu_1 - z & \mu_2 \\ \mu_3 & \mu_4 - z \end{vmatrix} \qquad \dots \quad (5.2)$$

Then (5.1) becomes

$$\begin{Vmatrix} \mu_1 - z & \mu_2 \\ \mu_3 & \mu_4 - z \end{Vmatrix} \begin{Vmatrix} \hat{N}_{1,n} \\ \hat{N}_{2,n} \end{Vmatrix} = \frac{z\rho}{D_1'} \begin{Vmatrix} \overline{N}_{1,0} \\ \overline{N}_{2,0} \end{Vmatrix}$$

and therefore

$$\hat{N}_{1,n} = \frac{z\rho}{D_1'\Delta} [\overline{N}_{1,0}(\mu_4 - z) - \overline{N}_{2,0}\, \mu_2] \qquad (5.3a)$$

$$\hat{N}_{2,n} = \frac{z\rho}{D_1'\Delta} [-\overline{N}_{1,0}\, \mu_3 + \overline{N}_{2,0}(\mu_1 - z)]. \qquad \dots \quad (5.3b)$$

Determinant Δ is given by equation (5.2) and is the same as (4.12). After fulfilling the stability condition (see (4.15)) this determinant has its absolute term equal to 0, i.e. $\lim_{\rho \to 0} \Delta(\rho) = 0$. Therefore the ratio $\frac{z\rho}{\Delta}$ can be cancelled by ρ and thus the degree

of the polynomial (4.12) decreases by one (the absolute term of this polynomial is, after the stability condition is satisfied, equal to zero !). Thus, we get the polynomial

$$\frac{\Delta}{\rho} = P(\rho).$$

By Tauber's theorem we get from (5.3a and 5.3b) the following expressions for limit values $\lim_{t\to\infty} \overline{N}_{1,n}(t)$, $\lim \overline{N}_{2,n}(t)$ which will be further denoted by $\overline{N}_{1,n,s}$ and $\overline{N}_{2,n,s}$

$$\overline{N}_{1,n,s} = \frac{z}{D_1'P(0)}\,[\overline{N}_{1,0}(\mu_{4,0}-z)-\overline{N}_{2,0}\,\mu_{2,0}] \qquad \dots \quad (5.4a)$$

$$\overline{N}_{2,n,s} = \frac{z}{D_1'P(0)}\,[-\overline{N}_{1,0}\,\mu_{3,0}+\overline{N}_{2,0}(\mu_{1,0}-z)] \qquad \dots \quad (5.4b)$$

In these equations $\mu_{1,0}$, $\mu_{2,0}$, $\mu_{3,0}$, $\mu_{4,0}$ are values of μ_1, μ_2, μ_3, μ_4 for $\rho = 0$. Clearly, by (4.11) $\mu_{1,0} = x_{1,0}$, $\mu_{2,0} = x_{2,0}$, $\mu_{3,0} = x_{3,0}$, $\mu_{4,0} = x_{4,0}$ and $P(0)$ is given by equation (4.14).

Thus we have everything prepared for the computation of stable values $\overline{N}_{1,n,s}$, and $\overline{N}_{2,n,s}$. It remains to consider the simplification in the case of equal basins. To do this it is sufficient to determine the interpretation of the symbols μ_1, μ_2, μ_3, μ_4, and especially of the determinant Δ (cf. (5.1) and (5.2)). As already said, determinant $v(\rho)$ is the determinant of the n-th power of the basic matrix, i.e. $|v(\rho)| = |a^n|$. According to Sylverster's theorem the power a^n can be expressed in eigenvalues λ_1, λ_2 and in the elements $\bar{a}+\rho$, $\bar{b}$, $\bar{c}$, $\bar{d}+\rho$ of the basic matrix a, so that

$$a^n = \beta_n \begin{Vmatrix} \bar{a}+\rho-\alpha_n, & \bar{b} \\ \bar{c}, & \bar{d}+\rho-\alpha_n \end{Vmatrix} \qquad \dots \quad (5.5)$$

where $$\alpha_n = \lambda_1\lambda_2\frac{\lambda_1^{n-1}-\lambda_2^{n-1}}{\lambda_1^n-\lambda_2^n},\ \beta_n = \frac{\lambda_1^n-\lambda_2^n}{\lambda_1-\lambda_2}.$$

It can be easily verified that matrix a^n has its eigenvalues equal to λ_1^n, λ_2^n. Therefore determinant Δ_n corresponding to matrix a^n-zj is given by the expression

$$\Delta_n = \beta_n \begin{vmatrix} \bar{a}+\rho-\alpha_n-z, & \bar{b} \\ \bar{c}, & \bar{d}+\rho-\alpha_n-z \end{vmatrix} \qquad \dots \quad (5.6)$$

Hence, we have $\mu_1 = \beta_n(\bar{a}+\rho-\alpha_n)$, $\mu_2 = \bar{b}\beta_n$, $\mu_3 = \bar{c}\beta_n$, $\mu_4 = \beta_n(\bar{d}+\rho-\alpha_n)$. The corresponding values for $\rho = 0$ are

$$\mu_{1,0} = \beta_{n,0}(\bar{a}-\alpha_{n,0}) \qquad \dots \quad (5.7a)$$

$$\mu_{2,0} = \beta_{n,0}\ \ b \qquad \dots \quad (5.7b)$$

$$\mu_{3,0} = \beta_{n,0}\ \ \bar{c} \qquad \dots \quad (5.7c)$$

$$\mu_{4,0} = \beta_{n,0}\ (\bar{d}-\alpha_{n,0}) \qquad \dots \quad (5.7d)$$

where α_n, β_n are also calculated for $\rho = 0$ and therefore denoted by $\alpha_{n,0}$, and $\beta_{n,0}$.

Finally, it remains to determine $P(0)$. In accordance with (4.6), for equal basins we have

$$v(\rho) = (a_0+\rho j)^n$$

and therefore

$$v(\rho)-zj = a_0^n+n\rho a_0^{n-1}+\ldots-zj.$$

Since $a_0 = \lim\limits_{\rho\to 0} a$ we have $a_0^n = \lim\limits_{\rho\to 0} a^n$ and similarly $a_0^{n-1} = \lim\limits_{\rho\to 0} a$. Then, we have

$$|\Delta_n = a_0^n+n\rho a_0^{n-1}+\ldots-zj| \qquad \ldots \quad (5.8)$$

where

$$a_0^n = \beta_{n,0}\begin{Vmatrix} \bar{a}-\alpha_{n,0}, & b \\ \bar{c}, & \bar{d}-\alpha_{n,0} \end{Vmatrix} \qquad \ldots \quad (5.9)$$

$$a_0^{n-1} = \beta_{n-1,0}\begin{Vmatrix} \bar{a}-\alpha_{n-1,0}, & b \\ \bar{c} \quad , & \bar{d}-\alpha_{n-1,0} \end{Vmatrix} \qquad \ldots \quad (5.10)$$

Although the coefficient at the linear term of determinant Δ_n can be easily computed, the general solution involves complicated expressions. Therefore we shall not deal with the general solution any further and will later give a special method for obtaining the solution in the case of a chain composed from equal basins.

Remark : More lucid results can be obtained when the eigenvalues of the matrix a^n-zj are used. These eigenvalues are given by $\Lambda_1 = \lambda_1^n-z$ and $\Lambda_2 = \lambda_2^n-z$, where λ_1 and λ_2 are the eigenvalues of matrix a (cf. (4.4)).

Consequently, determinant Δ_n of matrix a^n-zj is given by

$$\Delta_n = \Lambda_1\Lambda_2 = (\lambda_1^n-z)(\lambda_2^n-z) \qquad \ldots \quad (5.11)$$

Under the stability condition,

we have

$$\lim_{\rho\to 0} \Delta_n = 0$$

because

$$(\lambda_{1,0}^n-z)(\lambda_{2,0}^n-z) = 0 \quad \text{(cf.(4.16a) and (4.16b))}.$$

Assuming that by (4.4) $\quad \lambda_1 = \rho+\lambda_{1,0}, \qquad \lambda_2 = \rho+\lambda_{2,0}$

we have

$$\Delta_n = [(\rho+\lambda_{1,0})^n-z][(\rho+\lambda_{2,0})^n-z].$$

For determination of the coefficient at the linear term of $P(0)$ it suffices to consider the expression

$$(\lambda_{1,0}^n+n\rho\lambda_{1,0}^{n-1}-z)\ (\lambda_{2,0}^n+n\rho\lambda_{2,0}^{n-1}-z)$$

so that

$$P(0) = n(\lambda_{2,0}^n-z)\lambda_{1,0}^{n-1}+n(\lambda_{1,0}^n-z)\lambda_{2,0}^{n-1}$$

$$= n\lambda_{1,0}^{n-1}\lambda_{2,0}^{n-1}\Big[\lambda_{1,0}+\lambda_{2,0}-z\Big(\frac{1}{\lambda_{1,0}^{n-1}}+\frac{1}{\lambda_{2,0}^{n-1}}\Big)\Big].$$

Finally, we have

$$P(0) = n\lambda_{1,0}^{n-1}\lambda_{2,0}^{n-1}\Big[\lambda_{1,0}+\lambda_{2,0}-z\Big(\frac{1}{\lambda_{1,0}^{n-1}}+\frac{1}{\lambda_{2,0}^{n-1}}\Big)\Big]. \qquad \ldots \quad (5.12)$$

Thus $P(0)$ is determined and the result must coincide with that obtained from equations (5.8) through (5.10).

In order to demonstrate the introduced computation procedure and to show its numerical feasibility, several examples follow.

6. Numerical examples

1. The following values are given for cultivation in a single basin furnished with the circulation of the medium (Fig. 4 on p. 63) and with the take-off of the yield.

per day	per day
$\mu_{2,2} = 4$	$\delta_1 = 0,4$
$\mu_{2,4} = 2$	$\delta_2 = 2$
$\mu_{1,2} = 4$	$D'' = 3(D''_a = 2,\ D''_b = D' = 1)$.

By (2.1) we can easily verify that all parameters have dimension t^{-1}. In our case the unity is 1 day. The values were chosen in such a way that the stability condition (2.5) is satisfied. Our problem is to study transient phenomena and stable values for the average number of microorganisms in the basin and to compute the yield.

Although it is intuitively clear that the fourterminal in question is in fact "isolated" and thus equation (2.4) applies, we shall start from a more general equation (2.3). According to it we have after reaching given values

$$\left\| \begin{matrix} \hat{N}'_1 \\ \hat{N}'_2 \end{matrix} \right\| = \frac{1}{1} \left\| \begin{matrix} \rho+7,4; & -16 \\ -4 & ; \rho+11 \end{matrix} \right\| \left\| \begin{matrix} \hat{N}_1 \\ \hat{N}_2 \end{matrix} \right\| - \frac{\rho}{1} \left\| \begin{matrix} \bar{N}_{1,0} \\ \bar{N}_{2,0} \end{matrix} \right\|$$

where $\bar{N}_{1,0}$ and $\bar{N}_{2,0}$ are expected values for $t = 0$. We shall deal with them for a while as with general numbers. Clearly $\hat{N}'_1 = \hat{N}_1$ and $\hat{N}'_2 = \hat{N}_2$; thus we have

$$\text{(A)} \qquad \left\| \begin{matrix} \rho+6,4; & -16 \\ -4 & ; \rho+10 \end{matrix} \right\| \left\| \begin{matrix} \hat{N}_1 \\ \hat{N}_2 \end{matrix} \right\| = \rho \left\| \begin{matrix} \bar{N}_{1,0} \\ \bar{N}_{2,0} \end{matrix} \right\|$$

The determinant of the square matrix in (A) equals 0 for $\rho = 0$; this proves that the culture stability condition is fulfilled.

For $\hat{N}_1$ and $\hat{N}_2$ we get (A) from

$$\text{(B)} \qquad \hat{N}_1 = \frac{1}{\rho+16,4} [(\rho+10)\bar{N}_{1,0}+16\bar{N}_{2,0}]$$

$$\text{(C)} \qquad \hat{N}_2 = \frac{1}{\rho+16,4} [4\bar{N}_{1,0}+(\rho+6,4)\bar{N}_{2,0}]$$

These functions are Laplace-Wagner transforms of expected values $\bar{N}_1(t)$ and $\bar{N}_2(t)$ of the number of both studied types of microorganisms in dependence on time. Passing to limits in (B) and (C) for $\rho \to 0$ we get by Tauber's theorem the values of $\bar{N}_1(t)$ and $\bar{N}_2(t)$ for $t \to \infty$. Evidently

$$\text{(D)} \qquad \bar{N}_{1,s} = 0,61\bar{N}_{1,0}+0,975\bar{N}_{2,0}$$

$$\text{(E)} \qquad \bar{N}_{2,s} = 0,224\bar{N}_{1,0}+0,39\bar{N}_{2,0}$$

Clearly the asymptotic values $\bar{N}_{1,s}$ and $\bar{N}_{2,s}$ are linear combinations of the initial values $\bar{N}_{1,0}$ and $\bar{N}_{2,0}$, the ratio $\frac{\bar{N}_{1,s}}{\bar{N}_{2,s}}$ being equal to 2,5. A particular case is if

$$\bar{N}_{1,s} = \bar{N}_{1,0} \; ; \qquad \bar{N}_{2,s} = \bar{N}_{2,0}.$$

Then also the initial values must satisfy the relation

$$\text{(F)} \qquad \bar{N}_{1,0} = 2{,}5\bar{N}_{2,0}$$

If $\qquad \bar{N}_{1,0} < 2{,}5\,\bar{N}_{2,0}$ then $\bar{N}_{1,s} > \bar{N}_{1,0}$, $\bar{N}_{2,s} < \bar{N}_{2,0}$,

and if $\qquad \bar{N}_{1.0} > 2{,}5\,\bar{N}_{2.0}$ then $\bar{N}_{1.s} < \bar{N}_{1.0}$, $\bar{N}_{2,s} > \bar{N}_{2,0}$. (see Fig. 5 on p. 63)

Now let us find the inverse images of (B) and (C).

We get $\quad$ (G) $\quad \bar{N}_1(t) = 0{,}61\bar{N}_{1,0}+0{,}975\bar{N}_{2.0}+(0{,}39\bar{N}_{1,0}-0{,}975\bar{N}_{2,0})e^{-16,1t}$

(H) $\quad \bar{N}_2(t) = 0{,}244\bar{N}_{1,0}+0{,}39\bar{N}_{2.0}+(-0{,}244\bar{N}_{1,0}+0{,}61\bar{N}_{2.0})e^{-16.1t}$.

From these relations it follows that the culture process roughly stabilizes after the period

$$t_0 = \frac{3}{16,1} = 0{,}186 \sim \frac{1}{5} \text{ of day.}$$

Further, if $\bar{N}_{1,0} = 2{,}5\bar{N}_{2.0}$ then the culture process is stabilized immediately from the instant $t = 0$, i.e. no transient phenomena occur. Finally, by (3.5), the number of microorganisms in the yield reservoir is increasing and given by

$$\text{(K)} \quad \bar{N}_{1,R}(t)+\bar{N}_{2,R}(t) = \bar{N}_{TOT,\,R}(t) = D''_a \int_0^t (\bar{N}_1(\tau)+\bar{N}_2(\tau)dt$$

$$= (1{,}708\,\bar{N}_{1,0}+2{,}73\bar{N}_{2,0})t+(0{,}018\bar{N}_{1,0}-0{,}0454\bar{N}_{2,0})(1-e^{-16,1t}).$$

If $\bar{N}_{1,0} = 2{,}5\bar{N}_{2,0}$, the yield in the reservoir will increase linearly in time from the very beginning. Now, let us only consider the yield in the stable state. Assuming, we know $\bar{N}_{1,0}+\bar{N}_{2,0} = \bar{N}_{0,0}$ the slope of the growth of microorganisms in the yield reservoir is

$$1{,}708\,\bar{N}_{1,0}+2{,}73\bar{N}_{2,0} = 2{,}73N_{0,0}-1{,}022\,\bar{N}_{1,0}.$$

Evidently, this slope and therefore the yield also is maximal if $\bar{N}_{1,0} = 0$. This means, that at the beginning only mature microorganisms of the second type should be inset. In this model the magnitude of initial values does influence the yield. The yield is greater if the initial amount $N_{0,0}$ of microorganisms is greater. Assuming the same concentration it results in larger basin. For maximal output we must have

$N_{0,0} = \overline{N}_{2,0}$, $\overline{N}_{1.0} = 0$, so that in the stable state the amount of microorganisms in the yield reservoir increases according to the law

$$\overline{N}_{TOT,R} = 2,73\, N_{0,0} t$$

the ratio $\frac{\overline{N}_{1,s}}{\overline{N}_{2,s}}$ in the yield reservoir being equal to $0,975/0,39 = 2,5$. Thus, immature organisms prevail against mature ones.

2. As the second example we shall give the computation of continual culture of microorganisms in a closed system composed from three basins, in all of them the generation parameters and conditions are the same. The problem is to determine D_1' so that the stability condition is fulfilled, to compute stable values $\overline{N}_{1,s}$, $\overline{N}_{2,s}$ in all basins, and to determine the yield. Finally, we will compare this case with the subsequent one assuming a single basin and discuss the results obtained.

Concerning the parameters we assume in all three basins :

$$\mu_{2,2} = 4; \quad \mu_{2,4} = 2; \quad \mu_{1,2} = 4;$$
$$\delta_1 = 0,4; \quad \delta_2 = 2.$$

Further, we have

$$D_1'' = D_2'' = D_3'' = 3$$
$$D_2' = D_3' = 3.$$

The basic square matrix a reads

$$\text{(a)} \quad a = \begin{Vmatrix} \rho+7,4\,; & -16 \\ -4\,; & \rho+11 \end{Vmatrix}$$

Its characteristic values $(\lambda_0)_{1,2}$ for $\rho = 0$ follow from the equation

$$\lambda_0^2 - 18,4\lambda_0 + 17,4 = 0$$

and therefore

$$\text{(b)} \quad (\lambda_0)_{1,2} = \begin{cases} 17,4 \\ 1,0 \end{cases}$$

Since only the root $\lambda_{2,0}$ is admissible, the stability condition follows from equation (4.16b) and we have

$$D_1' D_2' D_3' = \lambda_{2,0}^3$$

i.e.

$$9D_1' = 1$$

so that

$$\text{(c)} \quad D_1' = \frac{1}{9} = 0,111 \qquad (D_1' < D_3'' = 3; \quad D_{a,3}'' = 2,889).$$

(If the root $\lambda_{1,0}$ and therefore equation (4.16a) is considered, then $D_1' > D_3''$, what contradicts our requirements).

Now, we shall turn our attention to the calculation of the stable state.

The stable amounts of microorganisms in the third basin are given by the formulae (5.6) and (5.7a-5.7d). First, we determine necessary parameters. We have

$$\text{(d)} \quad z = D_1' \, D_2' \, D_3' = 1.$$

By (5.7a,b,c,d) we get

$$\text{(e)} \quad \mu_{1,0} = (17,\, 4^3-1)\, \frac{7,\,4}{16,4} - 17,\, 4\, \frac{17,\, 4^2-1}{16,4} = 2048$$

$$\text{(f)} \quad \mu_{2,0} = (17,\, 4^3-1)\, \frac{-16}{16,4} = -5120$$

$$\text{(g)} \quad \mu_{3,0} = (17,4^3-1)\, \frac{-4}{16,4} = -1280$$

$$\text{(h)} \quad \mu_{4,0} = (17,4^3-1)\, \frac{11}{16,4} - (17,4^2-1)\, \frac{17,4}{16,4} = 3203$$

and by (5.12), (i) $\quad P(0) = 3.17,4^2 \left(18,4 - \dfrac{1}{17,4^2} - 1 \right)$

i.e. (j) $\quad P(0) = 15780.$

Hence, we can write

$$\overline{N}_{1,3,s} = \frac{9}{15780}\, [\overline{N}_{1,0}(3203-1) + 5120\, \overline{N}_{2,0}]$$

$$\overline{N}_{2,3,s} = \frac{9}{15780}\, [1280\, \overline{N}_{1,0} + \overline{N}_{2,0}(2048-1).$$

Thus (k) $\quad \overline{N}_{1,3,s} = 1,83\overline{N}_{1,0} + 2,92\overline{N}_{2,0}$

(l) $\quad \overline{N}_{2,3,s} = 0,73\overline{N}_{1,0} + 1,17\overline{N}_{2,0}.$

Substituting similarly as in the case of a single basin

$$\overline{N}_{1,0} + \overline{N}_{2,0} = N_{0,0} \quad \text{i.e.} \quad \overline{N}_{2,0} = N_{0,0} - \overline{N}_{1,0} \quad \text{into (k) and (l),}$$

we get (m) $\quad \overline{N}_{1,3,s} = 2,92\; N_{0,0} - 1,09\, \overline{N}_{1,0}$

(n) $\quad \overline{N}_{2,3s} = 1,17\, N_{0,0} - 0,44\, \overline{N}_{1,0}$

and we have

$$\frac{\overline{N}_{1,3,s}}{\overline{N}_{2,3,s}} = 2,5.$$

It is again evident, that the greatest yield is obtained if $\overline{N}_{1,0} = 0$ and $N_{0,0} = \overline{N}_{2,0}$. The yield increases in the stable state linearly according to the law

$$\text{(p)} \quad \overline{N}_{TOT,R} = D_a''(2,56\; \overline{N}_{1,0} + 4,09\overline{N}_{2,0})t.$$

Under optimal conditions, i.e. if $\overline{N}_{0,0} = N_{2,0}$ and if $D''_a = 2{,}889$ (cf.(c)), we have

$$\text{(q)} \quad \overline{N}_{TOT,R} = 11{,}8N_{00}t.$$

Taking into consideration that for a single basin we had $\overline{N}_{TOT,R} = 2{,}73N_{0,0}t$, we can conclude that for three basins and for the same $N_{0,0}$ the result is in absolute value cca 4 times greater. Now, let us compute the stable amounts of microorganisms in the first and the third basin.

Since

$$\left\| \begin{matrix} \overline{N}_{1,2,s} \\ \overline{N}_{2,2,s} \end{matrix} \right\| = \frac{a_0}{D'_3} \left\| \begin{matrix} \overline{N}_{1,3,s} \\ \overline{N}_{2,3,s} \end{matrix} \right\|$$

we have under optimal conditions $\overline{N}_{1,0} = 0$ in the second basin

$$\left\| \begin{matrix} \overline{N}_{1,2,s} \\ \overline{N}_{2,2,s} \end{matrix} \right\| = \frac{N_{0,0}}{3} \left\| \begin{matrix} 7{,}4\,; & -16 \\ -4; & 11 \end{matrix} \right\| \left\| \begin{matrix} 2{,}92 \\ 1{,}17 \end{matrix} \right\| = N_{0,0} \left\| \begin{matrix} 0{,}963 \\ 0{,}397 \end{matrix} \right\|$$

For the first basin we get

$$\left\| \begin{matrix} \overline{N}_{1,1,s} \\ \overline{N}_{2,1,s} \end{matrix} \right\| = \frac{N_{0,0}}{3} \left\| \begin{matrix} 7{,}4\,; & -16 \\ -4; & 11 \end{matrix} \right\| \left\| \begin{matrix} 0{,}963 \\ 0{,}397 \end{matrix} \right\| = N_{0,0} \left\| \begin{matrix} 0{,}2637 \\ 0{,}1717 \end{matrix} \right\|$$

For the stable state and optimal conditions we summarize as follows.

BASIN NUMBER	I	II	III
$\overline{N}_{1,i,s}/N_{0,0}$	0,2637	0,963	2,92
$\overline{N}_{2,i,s}/N_{0,0}$	0,1717	0,397	1,17

Assuming the same concentration C of alive microorganisms in all three basins, the total volume V_{TOT} of all basins in the second example is

$$\text{(s)} \quad V_{TOT} = \frac{N_{0,0}}{C}(2{,}92+0{,}963+0{,}2637+1{,}17+0{,}397+0{,}1717) = 5{,}88\frac{N_{0,0}}{C}$$

In the first example the volume of the single basin is

$$\text{(t)} \quad V = \frac{1{,}365}{C} N_{0,0}$$

Computing in both examples the yield W_1 per hour referred to the unity of the volume, we get for the second example (3 basins)

$$W_1 = \frac{11{,}8}{5{,}88} C = 2C$$

and for the first example (1 basin)

$$W'_1 = \frac{2{,}73}{1{,}365} C = 2C$$

The exploitation of the volume is the same here for both cases.

Example 3: Let us assume cultivation in two basins, provided in both of them are different conditions. Let for the first basin (see Fig. 1 on p. 59) $\mu_{1,0} = 0$; $\delta_1, \delta_2\, \mu_{2,2}, \mu_{2,4} \neq 0$ and, on the contrary, for the second basin $\mu_{2,2} = \mu_{2,4} = 0$ $\delta_1, \delta_2, \mu_{1,2} \neq 0$. Further, let D_1'' and D_2'' be given. The problem is whether and when such a system can be stable. Let us write the matrices $a_{1,0}$ and $a_{2,0}$ (see (4.5) and (4.5a)). With respect to the given values they are of the form

$$a_{1,0} = \begin{Vmatrix} \bar{a}_1 ; & -\bar{b}_1 \\ 0 ; & \bar{d}_1 \end{Vmatrix} \qquad a_{2,0} = \begin{Vmatrix} \bar{a}_2 ; & 0 \\ -\bar{c}_2 ; & \bar{d}_2 \end{Vmatrix}$$

Clearly, by (4.8)
$$v(0) = a_{1,0}\ a_{2,0} \begin{Vmatrix} \bar{a}_1, \bar{a}_2 + \bar{b}_1\bar{c}_2 ; & -\bar{b}_1\bar{d}_2 \\ -\bar{d}_1\bar{c}_2 ; & \bar{d}_1\bar{d}_2 \end{Vmatrix}$$

The trace of the matrix $v(0)$ is therefore $\bar{a}_1\bar{a}_2+\bar{b}_1\bar{c}_2+\bar{d}_1\bar{d}_2$ and the determinant is $a_1d_1a_2d_2$. The stability condition is by (4.13) equal to

$$z^2 - z(\bar{a}_1\bar{a}_2+\bar{b}_1\bar{c}_1+\bar{d}_1\bar{d}_2)+\bar{a}_1\bar{d}_1\bar{a}_2\bar{d}_2 = 0.$$

In the stable state this equation should have one root equal to zero. This is possible only under $\bar{a}_1\bar{d}_1\bar{a}_2\bar{d}_2 = 0$ what contradicts our assumptions. Near the stable state the term $\sqrt{\bar{a}_1\bar{d}_1\bar{a}_2\bar{d}_2}$ must be small compared with $\frac{1}{2}(\bar{a}_1\bar{a}_2+\bar{b}_1\bar{c}_2+\bar{d}_1\bar{d}_2)$. Then one root from $z_{1,2}$ is small and positive and the condition (4.16c) can be satisfied.

For detailed analysis we must remember that

$$\begin{array}{llll} \bar{a}_1 = \delta_{1,1}+D_1'' & \bar{b}_1 = 2(\mu_{2,2}+2\mu_{2,4}) & \bar{c}_1 = 0 & \bar{d}_1 = \delta_{2,1}+D_1''+\mu_{2,2}+\mu_{2,4} \\ \bar{a}_2 = \delta_{1,2}+D_2''+\mu_{1,2} & \bar{b}_2 = 0 & \bar{c}_2 = \mu_{1,2} & \bar{d}_2 = \delta_{2,2}+D_2'' \\ D_1'' = D_2' & z = D_1'D_2' = D_1'D_1'' & & 0 < D_1' < D_2'' \end{array}$$

Then one can see that to fulfil the stability condition is sometimes by no means simple.

7. Culture chain

Sometimes it is advantageous to couple several equal basins into a chain. The cultivation conditions are the same in all the links.

Though it is possible to compute chains using formulae given in Section 4, it seems purposeful, taking into account their specialities, to develop for them a special method of computation. To do this, we shall again start from equation (2.3) which holds for a single link, where we assume values $\overline{N}_{1,0}$ and $\overline{N}_{2,0}$ for $t = 0$. Computing the matrix products we get the equations

$$D'\hat{N}_{1,k-1} = a\hat{N}_{1,k} - b\hat{N}_{2,k} - \rho\overline{N}_{1,0} \qquad \dots \quad (7.1a)$$

$$D'\hat{N}_{2,k-1} = -c\hat{N}_{1,k} + d\hat{N}_{2,k} - \rho\overline{N}_{2,0} \qquad \dots \quad (7.1b)$$

We denoted

$$a = \rho+\delta_1+D''+\mu_{1,2} \qquad b = 2(\mu_{2,2}+2\mu_{2,4}) \qquad c = \mu_{1,2}$$

$$d = \rho+\delta_2+D''+\mu_{2,2}+\mu_{2,4};$$ k-varying index (see Fig. 7 on p. 63).

If all the links are the same, then also a, b, c, d, D' are the same. If the expected numbers of microorganisms at time $t = 0$ are the same in all links, then also the initial values $\overline{N}_{1,0}$ and $\overline{N}_{2,0}$ are the same for all the links and equations (7.1a) and (7.1b)

apply for every link of the chain. Therefore we can assume them as simultaneous differential equations describing the phenomena in the chain that had at the beginning the same concentration of microorganisms of both types in all the links. Similarly as in previous sections $\hat{N}_{1,k}$ and $\overline{N}_{2,k}$ are Laplace-Wagner transforms of the function $\overline{N}_{1,k}(t)$ and $\overline{N}_{2,k}(t)$, which determine expected values of the number of microorganisms of both types in dependence on time. To solve equations (7.1a, b) we first eliminate one of the unknowns, say $\hat{N}_{2,k}$. By (7.1a) we get

$$\hat{N}_{2,k} = \frac{1}{b}(a\hat{N}_{1,k} - D'\hat{N}_{1,k-1} - \rho\overline{N}_{1,0})$$

$$\hat{N}_{2,k-1} = \frac{1}{b}(a\hat{N}_{1,k-1} - D'\hat{N}_{1,k-2} - \rho\overline{N}_{1,0}) \quad \ldots \quad (7.2)$$

and substituting this result into (7.1b) we obtain after modification the equation

$$D'\hat{N}_{1,k}(a+d) - \hat{N}_{1,k+1}(ad-bc) - (D')^2\hat{N}_{1,k-1} = \rho\overline{N}_{1,0}(D'-d) - \rho b\overline{N}_{2,0} \quad \ldots \quad (7.3)$$

The particular solution of this equation is given by

$$\overline{N}_{1,k} = \text{const} = K_0 = \frac{\overline{N}_{1,0}(D'-d) - b\overline{N}_{2,0}}{D'(a+d) - (ad-bc) - (D')^2}. \quad \ldots \quad (7.4)$$

The solution of the homogeneous equation, corresponding to (7.3), reads

$$\hat{N}_{1,k} = e^{\gamma k}$$

where γ satisfies the characteristic equation

$$(D')^2 e^{-\gamma} - D'(a+d) + (ad-bc)e^{\gamma} = 0$$

i.e. the equation $\quad (D')^2 e^{-2\gamma} - D'(a+d)e^{-\gamma} + ad - bc = 0. \quad \ldots \quad (7.5)$

Since $a+d$ is the trace and $ad-bc$ the determinant of the basic matrix a, the eigenvalues of which we denote by λ_1, λ_2, we can rewrite equation (7.5) into

$$(D')^2 e^{-2\gamma} - D'(\lambda_1+\lambda_2)e^{-\gamma} + \lambda_1\lambda_2 = 0$$

which can be decomposed into the product

$$(D'e^{-\gamma} - \lambda_1)(De^{-\gamma} - \lambda_2) = 0$$

so that $\quad (e^{\gamma})_1 = \dfrac{D'}{\lambda_1} \qquad (e^{\gamma})_2 = \left(\dfrac{D'}{\lambda_2}\right).$

The general solution of the homogenized equation is therefore given by

$$\hat{N}_{1,k} = K_1\left(\frac{D'}{\lambda_1}\right)^k + K_2\left(\frac{D'}{\lambda_2}\right)^k$$

and the general solution of the equation (7.3) is given by

$$\hat{N}_{1,k} = K_1\left(\frac{D'}{\lambda_1}\right)^k + K_2\left(\frac{D'}{\lambda_2}\right)^k + \rho K_0. \quad \ldots \quad (7.6)$$

Now, we can derive the expression for $\hat{N}_{2,k}$. Using (7.6) we get by (7.2) after modification

$$\hat{N}_{2,k} = \frac{1}{b}\left[K_1\left(\frac{D'}{\lambda_1}\right)^k (a-\lambda_1) + K_2\left(\frac{D'}{\lambda_2}\right)^k (a-\lambda_2) + \rho(K_0(a-D') - \overline{N}_{1,0})\right] \quad \ldots \quad (7.7)$$

Expression (7.4) for K_0 can be simplified by introducing the characteristic value λ_1, λ_2 of the basic matrix. We can easily derive that

$$K_0 = \frac{b\overline{N}_{2,0}-\overline{N}_{1,0}(D'-d)}{(D'-\lambda_1)(D'-\lambda_2)}. \qquad \dots \quad (7.8)$$

The "constants" K_1, K_2 can be determined from the boundary conditions of the chain. In general, they as well as λ_1 λ_2, (see (4.4)), depend on ρ.

Let us compute the constants K_1, K_2 e.g. in the case that $\hat{N}_1$ and $\hat{N}_2$ are given at the beginning of the chain, provided $\overline{N}_{1,0}$ and $\overline{N}_{2,0}$ are equal to 0, i.e. in the moment $t = 0$ no microorganisms were present in the whole chain. We have by (7.6) and (7.7)

$$\hat{N}_{1\,0} = K_1+K_2$$

$$b\hat{N}_{2,0} = K_1(a-\lambda_1)+K_2(a-\lambda_2)$$

so that

$$K_1 = \frac{1}{\lambda_1-\lambda_2}\,[\hat{N}_{1,0}(a-\lambda_2)-b\hat{N}_{2,0}] \qquad \dots \quad (7.9)$$

$$K_2 = \frac{1}{\lambda_1-\lambda_2}\,[-\hat{N}_{1,0}(a-\lambda_1)+b\hat{N}_{2,0}] \qquad \dots \quad (7.10)$$

Substituting these expressions into (7.6) and (7.7), where we set, according to our assumption, initial values equal to zero, we get

$$\hat{N}_{1,k} = \frac{1}{\lambda_1-\lambda_2}\Big\{\hat{N}_{1,0}\Big[(a-\lambda_2)\Big(\frac{D'}{\lambda_1}\Big)^k-(a-\lambda_1)\Big(\frac{D'}{\lambda_2}\Big)^k\Big]$$
$$-\hat{N}_{2,0}\,b\Big[\Big(\frac{D'}{\lambda_1}\Big)^k-\Big(\frac{D'}{\lambda_2}\Big)^k\Big]\Big\} \qquad \dots \quad (7.11)$$

$$\hat{N}_{2,k} = \frac{-1}{\lambda_1-\lambda_2}\Big\{\hat{N}_{1,0}C\Big[\Big(\frac{D'}{\lambda_1}\Big)^k-\Big(\frac{D'}{\lambda_2}\Big)^k\Big]$$
$$+\hat{N}_{2,0}\Big[(a-\lambda_1)\Big(\frac{D'}{\lambda_2}\Big)^k-(a-\lambda_2)\Big(\frac{D'}{\lambda_1}\Big)^k\Big]\Big\} \qquad \dots \quad (7.12)$$

It can be easily checked, that for $k = 0$ the introduced conditions at the beginning of the chain are satisfied.

Remark : The results (7.11) and (7.12) can be modified using the substitution

$$\frac{D'}{\lambda_1} = Ce^{-\beta}\,\frac{D'}{\lambda_2} = Ce^{\beta}$$

Clearly

$$C = \frac{D'}{\sqrt{\lambda_1\lambda_2}} = \frac{D'}{\sqrt{ad-bc}}$$

$$\sin\,\beta = \frac{c(\lambda_1-\lambda_2)}{2D'}\cos\,\beta = \frac{c(\lambda_1+\lambda_2)}{2D'}$$

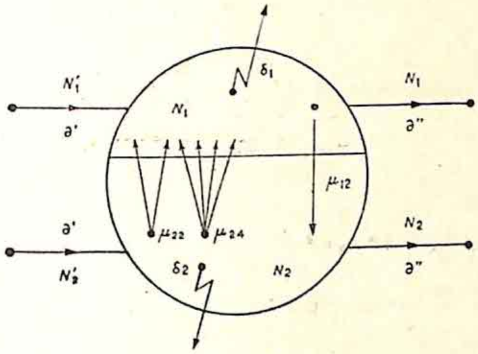

Fig. 1

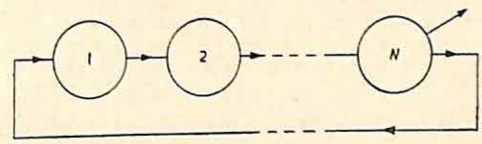

Fig. 2

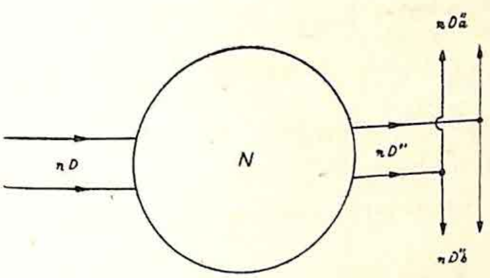

Fig. 3

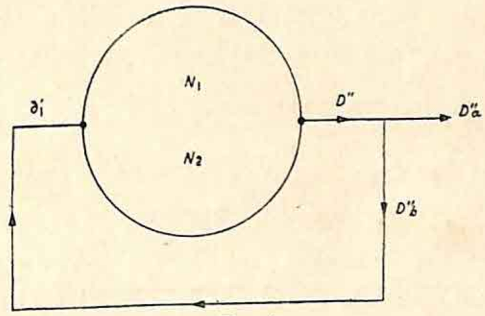

Fig. 4

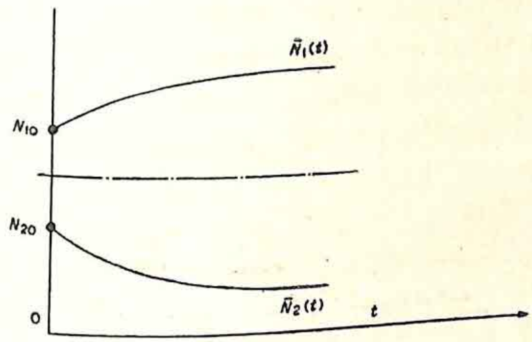

Fig. 5

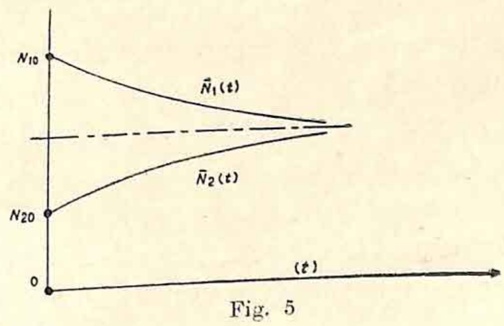

Fig. 5

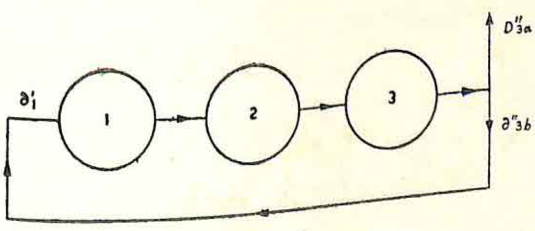

Fig. 6

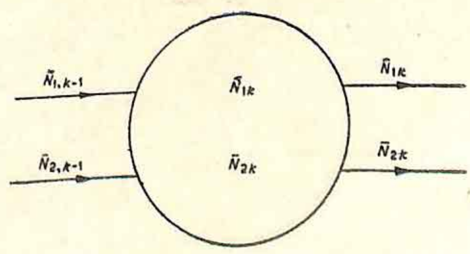

Fig. 7

and for C to be real we must have $ad > bc$. Equations (7.11) and (7.12) express the relation between Laplace-Wagner transforms of the expressions for expected values $\overline{N}_{1,k}(t)$ and $\overline{N}_{2,k}(t)$ at the end and at the arbitrary place of the chain under zero initial conditions. They can be also utilized for computing the stable state, when the expected values are independent of time. Then, it suffices to set everywhere $\rho \to 0$ and instead of λ_1, λ_2 use $\lambda_{1,0}$ $\lambda_{2,0}$,

where $$\lambda_{1,0} = \lim_{\rho \to 0} \lambda_1 \qquad \lambda_{2,0} = \lim_{\rho \to 0} \lambda_2 \quad \text{(see (4.4))}$$

The fact that $(D')^k$ is contained in all the terms, proves its significance. If $\lambda_{1,0} \geqslant \lambda_{2,0}$ then for $k \geqslant 1$ the terms $\left(\frac{D'}{\lambda_{1,0}}\right)^k$ can be omitted with respect to the terms $\left(\frac{D'}{\lambda_{2,0}}\right)$ and the geometric increase of $\hat{N}_{1,k}$ and $\hat{N}_{2,k}$ along the chain is very clear.

The behaviour of the chain at the moment $t = 0$ can be also easily derived from equations (7.11) and (7.12). It suffices to pass to limits in (7.11) and (7.12) according to Abel for $\rho \to \infty$. If $\hat{N}_{1,0}$ and $\hat{N}_{2,0}$ are constants, then, in the first approximation $\overline{N}_{1,k}$ and $\overline{N}_{2,k}$ increase in time similarly as the function t^n. That means, the greater is the distance from the beginning of the chain (i.e. the greater k), the greater the delay. This results from the fact, that microorganisms stay for some time in every basin before they proceed further.

8. Conclusions

In the present paper some important questions of continuous cultivation of microorganisms (e.g. algae) are dealt with. This stochastic process is studied as a linear multiparametric branching process birth, aging, migration, and death. The purpose of the paper is to compute expected values and variances of the number of particular types of microorganisms in cultivation basins with continual circulation of the medium which is supplied with nourishing substances and from which some part of microorganisms is taken off as the yield. The conditions in the whole of the basin are assumed to be the same. In the paper the formulae both for the stationary and for the non-stationary state as well as the stability condition are given. A special attention is paid to the case, when the system consists of several equal basins, i.e. when the system represents a chain. Besides general formulae, which enable the computation of the model, several numerical examples are given.

It follows from calculations, that in the steady state there is no essential qualitative difference between the cultivation in a single basin and in a system of mutually coupled basins. But the transient phenomena are different.

References

Bellman, R., Harris, T. E. and Bharucha-Reid, A. T. (1960) : *Elements of the Theory of Markov Process*, McGraw-Hill, 96.

Feller, W. (1958) : *An Introduction to Probability Theory*, **1**, 411, John Wiley & Sons, New York.

Pol, B. V. D. and Bremer, H. (1950) : Tauber's Theorem, *Operational Calculus*, Cambridge.

Paper received : January, 1963.

THEORY OF AN EXPERIMENT ON THE REPLICATION OF DNA

By P. A. P. MORAN

Australian National University, Canberra

SUMMARY. The change in composition of bacterial DNA in a bacterium is calculated when the bacterium is transferred to a medium containing a different isotope of nitrogen and certain assumptions made about the mode of replication of the chromosome. The results can be compared with experimental measurements.

Meselson and Stahl (1958) determined the change in composition of bacterial DNA following a change in medium from one containing all its nitrogen as N^{15} to one containing only N^{14}. They showed that the DNA first became a hybrid, containing equal amounts of N^{15} and N^{14}, and then became a mixture of hybrid and pure N^{14} DNA. These experiments therefore confirmed the predictions of Watson and Crick (1953) that DNA replicates "semiconservatively", each strand of the double helix being associated with a strand of new material.

The units of DNA studied in these experiments each comprised about one hundredth of the bacterial chromosome, and inside the bacterium are probably joined end-to-end . Since few, if any, pure N^{14} units appeared until most of the units were $N^{15}N^{14}$ hybrids it is clear that they are not replicated at random. Their behaviour appears to fit a scheme in which the whole chromosome is replicated by one process starting at one end and moving progressively to the other.

For convenience we shall refer to the $N^{15}N^{15}$, $N^{15}N^{14}$, and $N^{14}N^{14}$ units as "fully labelled", "half-labelled", and "unlabelled" and we assume that the relative proportions of these units are the relative proportions of the lengths of chromosomes which are $N^{15}N^{15}$, $N^{15}N^{14}$, and $N^{14}N^{14}$.

The purpose of the present paper is to calculate these proportions on one theory of the manner in which the chromosome replicates. The length of chromosome is taken as unity and the time scale is chosen so that the population doubles in unit time.

The chromosomes are assumed to consist of two strands joined along their length which are torn apart starting from one end, a replicate being immediately formed at the point of division, this replicate containing nitrogen molecules from the medium. From the above conventions about the units of time and length, the velocity of the point of division along the arm is also unity.

When the process of tearing is complete there will be two complete chromosomes, each double-stranded, where there was only one before. It is then assumed that the two ends of each newly formed chromosome join together forming a chromosomal ring, and that this ring is then broken at a point which is randomly and uniformly distributed about the circumference. The process of division then immediately starts again at one of the two ends.

Let n be the number of chromosomes in all the cells at time $t = 0$, and $n(t)$ the number at time t. Then

$$n(t) = ne^{\beta t} \qquad \dots \quad (1)$$

where $\beta = ln2 = 0.6931$. Let $N(t)$ be the total length of all the arms of the chromosomes so that if the length of tearing of an individual chromosome is x, the total length of the arms in this chromosome will be $1+x$. $N(t)$ will be a constant multiple of $n(t)$ and if N is its initial value

$$N(t) = Ne^{\beta t}. \qquad \dots \quad (2)$$

We must first determine the probability distribution of x. It is clear that this has a continuous probability density between 0 and 1 with no concentrations of probability at any point. Let this probability density be $f(x)$. The number of chromosomes at time T with x's between x and $x+dx$ will be $(0 < x < 1-dx)$

$$n(T)f(x)dx. \qquad \dots \quad (3)$$

Consider the corresponding frequency at a time $T+\Delta T$ where ΔT is small but dx much smaller, and we also assume $x < 1-\Delta T$. Then this frequency is

$$n(T)e^{\beta \Delta T} f(x)dx,$$

and is also equal to

$$n(T)f(x-\Delta T)dx.$$

Thus

$$f(x-\Delta T) = f(x)e^{\beta \Delta T},$$

and letting $\Delta T \to 0$ we conclude that $f(x)$ is proportional to

$$e^{-\beta x},$$

so that $f(1) = \frac{1}{2}f(0)$ which checks with the fact that there are obviously just twice as many chromosomes beginning to split as ones which are nearly completely split. Since the integral of $f(x)$ over $(0, 1)$ must be unity we have

$$f(x) = 2\beta e^{-\beta x}$$
$$= 2^{1-x}\, ln2. \qquad \dots \quad (4)$$

Let $A(t)$, $B(t)$, $C(t)$ be the total length of those parts of the arms of chromosomes which are respectively fully labelled, half-labelled, and unlabelled. We have

$$N(t) = A(t)+B(t)+C(t) \qquad \dots \quad (5)$$

and at $t = 0$,

$$A(0) = N, \quad B(0) = C(0) = 0. \qquad \dots \quad (6)$$

We calculate the change of $A(t)$, $B(t)$, and $C(t)$ with time, beginning with the interval $0 < t < 1$. A chromosome chosen at random will be split to a length x where x has the distribution (4). If $x > t$ the chromosome was originally formed at a time before zero and then consisted entirely of fully labelled parts. At time t, fully labelled material is being destroyed and replaced by twice as much half-labelled material. On the other hand if $x < t$, the chromosome was formed at time $t-x > 0$ and therefore then consisted of two portions, one of length $t-x$ of half-labelled material, and one of length $1-t+x$ of fully labelled material. Since the chromosome formed itself into a ring and then broke at a random point, the actual point of division at time t is such that fully labelled material is being split with probability $1-t+x$, and

half labelled material with probability $t-x$. Furthermore, at time t the total number of chromosomes is $ne^{\beta t}$. Putting these facts together we find that

$$A'(t) = -ne^{\beta t}\left\{\int_t^1 f(x)dx + \int_0^t (1-t+x)f(x)dx\right\}, \qquad \dots (7)$$

$$B'(t) = -2A'(t), \qquad \dots (8)$$

$$C'(t) = ne^{\beta t}\int_0^t (t-x)f(x)dx. \qquad \dots (9)$$

Considering (7) first and inserting (4) we get

$$A'(t) = -ne^{\beta t} + 2nte^{\beta t} + 2n\beta^{-1} - 2n\beta^{-1}e^{\beta t}.$$

Integrating this and using the fact that $A(0) = N$ we get

$$A(t) = n\beta^{-1}\{2-e^{\beta t}+2t\}+n\beta^{-2}\{4+2(\beta t-1)e^{\beta t}-2e^{\beta t}\}, \quad (0 \leqslant t \leqslant 1). \qquad \dots (10)$$

At $t = 1$ this is

$$A(1) = n\beta^{-1}(4-e^{\beta})+n\beta^{-2}\{4+2(\beta-2)e^{\beta}\}, \qquad \dots (11)$$

which can be simplified on using $e^{\beta} = 2$.

Similarly by integrating (8) and (9) and using the fact that $B(0) = C(0) = 0$, we find

$$B(t) = 2n\beta^{-1}\{e^{\beta t}-1-2t\}+4n\beta^{-2}\{(2-\beta t)e^{\beta t}-2\}, \qquad \dots (12)$$

$$B(1) = -2n\beta^{-1}(3+e^{\beta})+8n\beta^{-2}(e^{\beta}-1), \qquad \dots (13)$$

$$C(t) = 2n\beta^{-1}t+2n\beta^{-2}(2+(\beta t-2)e^{\beta t}), \qquad \dots (14)$$

$$C(1) = 2n\beta^{-1}(e^{\beta t}+1)+4n\beta^{-2}(1-e^{\beta}). \qquad \dots (15)$$

These check with the fact that we must have

$$A(t)+B(t)+C(t) = n\beta^{-1}e^{\beta t}. \qquad \dots (16)$$

Next consider the interval $1 \leqslant t \leqslant 2$. We again choose a chromosome at random and consider its x. If $x > t-1$ the chromosome was originally formed in the period $(0 \leqslant t \leqslant 1)$ and it was completed at the time $t-x < 1$. At that time it must therefore have consisted of a length $t-x$ of half-labelled material and a length $1-t+x$ of fully labelled material. After forming a ring it splits at a randomly chosen point and therefore at time t the part which is dividing is fully labelled with probability $1-t+x$ and half labelled with probability $t-x$.

On the other hand if $x < t-1$ the chromosome was originally completed at time $t-x > 1$. It therefore consists of a portion half-labelled of length $2-t+x$ and an unlabelled portion of length $t-x-1$. These are therefore the respective probabilities of the portions being torn at time t. Putting these facts together we therefore get

$$A'(t) = -ne^{\beta t}\left\{\int_{t-1}^1 (1-t+x)f(x)dx\right\}, \qquad \dots (17)$$

$$B'(t) = 2ne^{\beta t}\left\{\int_{t-1}^1 (1-t+x)f(x)dx\right\}, \qquad \dots (18)$$

$$C'(t) = ne^{\beta t}\left\{\int_{t-1}^1 (t-x)f(x)dx + \int_0^{t-1} f(x)fx\right\}. \qquad \dots (19)$$

Integrating these equations and using the initial values (11), (13) and (15) we get for $1 \leqslant t \leqslant 2$

$$\beta n^{-1}A(t) = 8-8\beta^{-1}-4t+e^{\beta t}\{2-t+2\beta^{-1}\}, \quad \ldots \quad (20)$$

$$\beta n^{-1}B(t) = -14+16\beta^{-1}+8t-2e^{\beta t}\{2-t+2\beta^{-1}\}, \quad \ldots \quad (21)$$

$$\beta n^{-1}C(t) = 6-8\beta^{-1}-4t+e^{\beta t}\{3-t+2\beta^{-1}\}. \quad \ldots \quad (22)$$

After $t = 2$, $A(t) = 0$, and $B(t)$ remains constant. The course of $A(t)$, $B(t)$, and $C(t)$ during the interval $0 \leqslant t \leqslant 2$ is given in the following table :

TABLE

t	$\beta n^{-1} A$	$\beta n^{-1} B$	$\beta n^{-1} C$
0	1.000	0	0
0.2	0.857	0.286	0.005
0.4	0.689	0.622	0.009
0.6	0.525	0.950	0.041
0.8	0.369	1.262	0.110
1.0	0.229	1.542	0.229
1.2	0.120	1.760	0.416
1.4	0.060	1.880	0.700
1.6	0.020	1.960	1.052
1.8	0.002	1.996	1.484
2.0	0	2.000	2.000

It is easy to check that $\beta n^{-1}\{A(t)+B(t)+C(t)\} = e^{\beta t}$. From measurements of the ultraviolet absorption bands of the DNA such as are given in Meselson and Stahl's paper it would be possible to compare the above results with experiments, and also the predictions of alternative theories. Thus we might suppose that as soon as a new chromosome is completed and separated from its twin it begins dividing at once starting from one end without first forming a circle. In such a case we might suppose that, (a) it begins to divide from the same end as in the previous division; (b) it begins to divide from the opposite end to the previous division; (c) it begins to divide from either end at random. The results for each of these cases will differ from each other, and from (20), (21), and (22), and can be easily worked out by the same kind of argument.

I am indebted to Dr. H. J. Cairns for informing me of this problem.

REFERENCES

MESELSON, M. and STAHL, F. W. (1958): The replication of DNA in Escherica Coli. *Proc. Nat. Acad. Sci.* USA, **44**, 671–682.

WATSON, J. D. and CRICK, F. H. C. (1953): Genetical implications of the structure of Deoxyribonucleic acid, 4361, *Nature*, **171**, 964.

Paper received : December, 1962.

ON FINITENESS OF THE PROCESS OF CLUSTERING*

By JERZY NEYMAN

Statistical Laboratory, University of California, Berkeley

SUMMARY. The study of the process of clustering leads to the consideration of the sum, say N, of random variables, where the number of components is random and unbounded. As a result, N may be infinite with probability one. In addition to three easy theorems on the subject, the paper gives two examples where the divergence of N to infinity occurs in somewhat unexpected conditions.

1. Introduction

It is known (Neyman, 1955), that the most general homogeneous first order process of clustering can be generated by the following mechanism, supposed to operate in the Euclidean space E of arbitrary dimensionality.

(i) Particles C, described as cluster centers, are Poisson-wise distributed in E. Let λ represent the expected number of cluster centers in any Borel subset of E of measure unity.

(ii) With every cluster center C there is associated a positive integer valued random variable ν, described as the number of members of the cluster centered at C. Numbers ν attached to different cluster centers are mutually independent and have the same distribution characterized by the probability generating function $G_\nu(t) = E(t^\nu)$, where $|t| \leqslant 1$.

(iii) Given the position of a cluster center C and given that the corresponding $\nu = n \geqslant 1$, n particles $\pi_1, \pi_2, \ldots, \pi_n$, described as members of the cluster centered at C, are distributed in space E in accordance with the following postulates. Let $u = (u_1, u_2, \ldots)$ represent the coordinates of C and $X^{(i)} = (X_1^{(i)}, X_2^{(i)}, \ldots)$ the coordinates of the i-th member of the cluster. It is postulated that $X^{(i)}$ is a random variable with probability density, to be denoted by $f(x-u)$ and called the "structure" of the cluster, depending only upon the distance between the cluster center C and the point $x = (x_1, x_2, \ldots)$ at which it is evaluated. The random variables $X^{(i)}$ for $i = 1, 2, \ldots, n$ are mutually independent. Also, they are independent of similar variables corresponding to other cluster centers.

(iv) For each particle π located at some point x there is performed a random trial capable of yielding a "success" or a "failure." $\theta(x)$ represents the probability of success, and we assume that it is a Borel measurable function of x. The "success" of one particle is independent of that of any other particle and also of all other variables in the system.

Let R stand for a Borel set in E having a positive, possibly an infinite, measure. The subject of our study is the random variable N representing the number of those particles π in R that are "successful." In particular, in Section 2 we study the conditions under which $P\{N = +\infty\} = 1$ in which case we shall say that N is degenerate. Sections 3 and 4 are given to examples of the degeneracy of N that appear to have certain unexpected features.

*Prepared with the partial support of the National Science Foundation, Grant GP—10.

2. Conditions for N to be nondegenerate

The probability generating function of N is given (Neyman and Scott, 1952) by the formula

$$G_N(t) = e^{-\lambda\int\{1-G_\nu[1-(1-t)p(u)]\}du}, \qquad \ldots \quad (2.1)$$

where the integral in the exponent extends over the whole space E and where

$$p(u) = \int_R f(x-u)\theta(x)dx \qquad (2.2)$$

represents the probability that a particle π belonging to a cluster centered at u will be in R and that it will be "successful."

The integral in (2.1) is never negative. If it is convergent, then (2.1) is equal to unity at $t = 1$ and N is finite with probability one. On the other hand, if for a fixed value of t between limits $0 \leqslant t < 1$ the same integral diverges to infinity, then $P\{N = k\} = 0$ for all finite k, and N is degenerate. It follows that to study the conditions under which N is degenerate means to study the conditions under which the integral in the right hand side of (2.1) is divergent.

Theorem 1 : *In order that N be nondegenerate, it is necessary that*

$$I = \int_E p(u)du < +\infty. \qquad \ldots \quad (2.3)$$

The proof is based on the following simple inequality. For any nonnegative $w \leqslant 1$

$$1-G_\nu(w) = 1-\sum_{n=0}^{\infty} w^n P\{\nu = n\} \geqslant 1-P\{\nu = 0\}-w(1-P\{\nu = 0\}), \qquad \ldots \quad (2.4)$$

or

$$1-G_\nu(w) \geqslant (1-P\{\nu = 0\})(1-w). \qquad \ldots \quad (2.5)$$

It follows that the integrand in (2.1)

$$1-G_\nu[1-(1-t)p(u)] \geqslant (1-P\{\nu = 0\})(1-t)p(u) \geqslant 0. \qquad \ldots \quad (2.6)$$

Thus, if the integral in (2.1) converges then I must be finite.

Theorem 2 : *If the number ν of particles in a cluster has a finite expectation*

$$\nu_1 = E(\nu) = G_\nu'(1), \qquad \ldots \quad (2.7)$$

then the convergence of I is a sufficient condition for nondegeneracy of N.

Theorem 2 is due to Mr. A. H. Marcus. For any positive $w < 1$ we have

$$1-G_\nu(w) = (1-w)G_\nu'(w^*), \qquad \ldots \quad (2.8)$$

where w^* is a number between w and unity. However, the derivative of a probability generating function is nondecreasing. It follows that

$$1-G_\nu(w) \leqslant (1-w)G_\nu'(1) = (1-w)\nu_1 \qquad \ldots \quad (2.9)$$

and, in particular

$$1-G_\nu[1-(1-t)p(u)] \leqslant \nu_1(1-t)p(u). \qquad \ldots \quad (2.10)$$

Thus, if I is finite then the integral in (2.1) is convergent and N is nondegenerate.

Theorem 3 : *If the expectation ν_1 and the measure of R are both finite, then N is nondegenerate.*

In order to prove this theorem it is sufficient to show that the hypotheses imply the finiteness of I. Using (2.3) we have

$$I = \int_E du \int_R f(x-u)\theta(x)dx, \qquad \ldots \quad (2.11)$$

and, changing the order of integration and remembering that f is a probability density, the integral of which over E is equal to unity,

$$I = \int_R \theta(x)dx \leqslant m(R). \qquad \ldots \quad (2.12)$$

Naturally, depending upon the nature of the probability $\theta(x)$, the integral I may be finite even if the region R has an infinite measure. For example, this is the case when the clusters considered are those of galaxies and "success" means sufficient apparent brightness for a galaxy to be included in a catalogue. Here R may mean the whole three dimensional Euclidean space, with infinite measure. However, $\theta(x)$ tends to zero with increasing distance fast enough for the integral (2.12) to converge.

The above circumstances may suggest that when I is finite, the supplementary assumption of Theorem 2, that the expectation ν_1 is also finite, does not play any role and that the finiteness of I by itself may guarantee the nondegeneracy of N. The example in the next section shows that this presumption is false.

3. Example of a bounded region R with a degenerate N

Let E be a straight line and R be the interval $[-1, +1]$. Also assume $\theta(x) \equiv 1$. Thus $I = 2$. We shall define a probability generating function $G_\nu(t)$ and the structure of clusters $f(x-u)$ such that the number N of particles in R will be infinite with probability one. Because of Theorem 3, this can be possible only if the expectation of ν is infinite.

For $n = 2, 3, \ldots$ let

$$P\{\nu = n\} = 1/n(n-1). \qquad \ldots \quad (3.1)$$

It is easy to see that the corresponding probability generating function is

$$G_\nu(t) = t+(1-t)\log(1-t) \quad \text{for} \quad 0 \leqslant t < 1. \qquad \ldots \quad (3.2)$$

Also, it is obvious that $t \to 1$ implies $G_\nu(t) \to 1$. Further,

$$1-G_\nu[1-(1-t)p(u)] = (1-t)[1-\log(1-t)]p(u)-(1-t)p(u)\log p(u). \qquad \ldots \quad (3.3)$$

The integral over E of $p(u)$ being equal to 2, irrespective of the cluster structure f, the degeneracy or nondegeneracy of N depends on whether the integral

$$J = \int_E |p(u)\, \log\, p(u)|\, du \qquad \ldots \quad (3.4)$$

diverges or converges, respectively. The answer to the latter question depends upon the behaviour of $p(u)$ for large values of $|u|$. Obviously, $p(u)$ is symmetrical with respect to the origin of coordinates. Therefore, it will be sufficient to consider only values of $u > a$ and it will be convenient to set $a > e$ so that $\log u > 1$. Our purpose is to find $p(u)$ such that

$$\int_a^\infty p(u)du < +\infty, \text{ while } -\int_a^\infty p(u)\log p(u)du = +\infty. \qquad \ldots \quad (3.5)$$

An example of such a function $p(u)$ is provided by

$$p(u) = \frac{C}{u \log^2 u} \qquad \ldots \quad (3.6)$$

where C is a constant to be used in defining f. However, the value of this constant does not affect the convergence of the integrals in (3.5).

Easy calculations show that, for any $a < b$,

$$\int_a^b \frac{du}{u \log^2 u} = \frac{1}{\log a} - \frac{1}{\log b}. \quad \ldots \quad (3.7)$$

Thus, if $p(u)$ is given by (3.6) then the first of the integrals in (3.6) has a finite value $C/\log a$. Similar calculations give

$$\int_a^b p(u)|\log p(u)|du = C \int_a^b \frac{du}{u \log^2 u} \, |\log C - \log u - 2 \log \log u| \quad \ldots \quad (3.8)$$

and it is seen that the convergence of the second integral in (3.5) depends on that of

$$\int_a^b \frac{\log u + 2 \log \log u}{u \log^2 u} \, du > \int_a^b \frac{du}{u \log u} = \log \log b - \log \log a. \quad \ldots \quad (3.9)$$

However, as $b \to \infty$, the last expression tends to infinity, and it follows that, with probability equal to one, the number N of particles in $[-1, +1]$ is infinite.

At this point the question may arise whether formula (3.6) is consistent with the meaning of $p(u)$ defined by (2.2) with $\theta(x) \equiv 1$. In other words, it is appropriate to ask whether a probability density f can be defined so that, at least for $|u| \geqslant a$,

$$\int_{-1}^{+1} f(x-u)dx = p(u), \quad \ldots \quad (3.10)$$

where $p(u)$ is given by (3.6). The answer is in the affirmative. In fact, for x between limits $|x| < a-1$ we may take any even nonnegative integrable function $\phi(x)$ and set $f(x) = C_1\phi(x)$, where C_1 is an adjustable constant. For $x > a-1$ the definition of f must conform with the properties of $p(u)$. We notice that for $u > a$, the function $p(u)$ is decreasing. Thus its derivative is negative. The definition of the corresponding $f(x)$ is made separately for each interval

$$a+2n-1 < x < a+2n+1 \quad \ldots \quad (3.11)$$

with $n = 0, 1, 2, \ldots$. Namely, for each n and $|\xi| < 1$,

$$f(a+2n+\xi) = -\sum_{m=1}^{\infty} p'[a+2(n+m)-1+\xi] \quad \ldots \quad (3.12)$$

and it is easy to check that the series on the right is convergent and positive. Also, it is easy to verify that (3.12) satisfies (3.10) for any $u > a$. Since f is an even function, its definition is now complete. This definition depends on two constants C and C_1 which can be adjusted easily so that the integral of f over E is equal to unity.

To appreciate the result obtained fully, consider its "operational" interpretation as follows. Let E denote a straight wire extending to infinity in both directions. Some moths are laying batches of eggs on this wire distributing them Poisson-wise with an arbitrarily law density, say one batch of eggs per mile, on the average. The number of eggs varies from one batch to the next, namely one-half of all batches contain two eggs, one-sixth of them three eggs, etc. Generally, the probability that a batch will contain exactly n eggs is $1/n(n-1)$. The eggs hatch, and the resulting larvae begin to crawl along the wire. They do so independently from each other. The probability that, after a certain time, a larva starting from a point u on E will be found at some close distance is more or less arbitrary. However, when we come to large

distances, the probability of the larva being found in a one-foot interval centered at a point x is given by (3.6) with $|x-u|$ replacing u. The question is about the probability that the number N of larvae in a one-foot length of the wire E, namely in the one foot-length centered at the origin of coordinates, will exceed one million. Unexpectedly to the present writer, the probability in question is equal to unity. The feeling of surprise concerning this result stems from the well-established habit of "approximating" infinite regions of space by finite but large sections. Thus, we are prone to consider that a long wire may well stand for an infinite wire E. Yet, if in the above example the infinite wire E is replaced by a finite wire extending from $-A$ to $+A$, then the integral (3.4) will have to be taken between the same limits and will be convergent so that the random variable N will be nondegenerate.

The question may be asked whether, with the distribution of ν defined by (3.2) and with an infinite E, there are cases where N is nondegenerate. The answer is in the affirmative. For example, if the structure of the cluster $f(x-u)$ vanishes whenever $|x-u|$ exceeds a certain limit, then the integrand in (2.1) differs from zero only over a set of points of finite measure and must converge no matter what the nature of $G_\nu(t)$ is. The general conclusion is that, if $E(\nu)=+\infty$, the nondegeneracy of N depends on the inter-relationship between the speed of decrease of $P\{\nu = n\}$ as $n\to\infty$ and the speed of convergence to zero of $p(u)$ as $|u|\to\infty$. Here a general theorem establishing the necessary and sufficient conditions for the nondegeneracy of N would be interesting.

4. Nondegeneracy of the number of images of galaxies on a photographic plate

Let E stand for the three dimensional Euclidean space. We treat the galaxies as dimensionless luminous particles distributed in E with clustering of first order. We assume that the expected number ν_1 of galaxies per cluster is finite. Also we assume that each galaxy emits photons in all directions Poisson-wise, their average number per unit time and per unit solid angle being μ^*, the same for all galaxies. The emissions by particular galaxies are mutually independent and the energy of all the photons is the same, independent of the distances of galaxies.

A photograph of a region R in the sky is taken with a telescope. The area of the telescope's mirror is A. The photons hitting the mirror are directed towards a point on the photographic plate depending upon the position of the galaxy emitting them. The exposure time is T units. Our basic assumption is that, in order that the image of a galaxy be visible on the photographic plate it is necessary and sufficient that the number of photons from this galaxy reaching the mirror of the telescope during the exposure time T be at least equal to a fixed number s. Let N denote the number of images of galaxies on the photographic plate. The problem is to find under what conditions, if any, the random variable N is degenerate.

It will be realized that the above assumptions regarding galaxies, about their luminosity properties, etc. are gross over-simplifications. However, this circumstance does not detract from the interest of the example considered.

It will be realized that the distribution of N is determined by formula (2.1) provided one defines appropriately the meaning of the "success" of any particular galaxy and calculates the probability $\theta(x)$ of this success, given that the coordinates of the galaxy in space are $x = (x_1, x_2, x_3)$.

Let x be a point within the solid angle R over which the photograph of the sky is taken. We assume that the origin of coordinates is at the observer and denote by ξ the distance of the point x,

$$\xi = \{x_1^2, +x_2^2+x_3^2\}^{\frac{1}{2}}. \qquad \dots \quad (4.1)$$

Obviously, in order that a galaxy located at x be "successful" in having its image recorded on the photographic plate, it is necessary and sufficient that during time T this galaxy emits at least s photons in the direction of the telescope's mirror. To a sufficient degree of approximation, the number of photons so emitted during time T is a Poisson variable, say Y, with expectation

$$E(Y) = \mu^* TA/\xi^2 = \mu/\xi^2, \quad \text{say}. \qquad \dots \quad (4.2)$$

Thus, the probability of "success" of the galaxy is

$$\theta(x) = P\{Y \geqslant s \,|\, x\} = e^{-\mu/\xi^2} \sum_{n=s}^{\infty} \frac{(\mu/\xi^2)^n}{n!} = \frac{1}{(s-1)!} \int_0^{\mu/\xi^2} e^{-t} t^{s-1} dt. \qquad \dots \quad (4.3)$$

According to Theorems 1 and 2, in order that N be nondegenerate, it is necessary and sufficient that the integral

$$I = \int p(u)du = \int_R \theta(x)dx = \int_R \frac{dx}{(s-1)!} \int_0^{\mu/\xi^2} t^{s-1} e^{-t}\, dt \qquad \dots \quad (4.4)$$

be convergent. R being a solid angle with its vertex at the origin of coordinates, the transformation to polar coordinates easily leads to the formula

$$I = C \int_0^{\infty} \xi^2 \int_0^{\mu/\xi^2} t^{s-1}\, e^{-t}\, dt, \qquad \dots \quad (4.5)$$

where C is a constant depending upon s and R. Now, it is easy to verify that the integral (4.5) converges when $s \geqslant 2$ and diverges when $s = 1$, irrespective of the value of $\mu = \mu^* TA > 0$. Thus, we come to the following paradoxical conclusion: the finiteness of the number of images of galaxies on the photograph of the sky depends only upon the sensitivity of the emulsion, but not on the size of the telescope and not on the length of exposure. If the emulsion of the photographic plate is so sensitive as to be able to record a single photon, then, with probability one, the number of images of galaxies in the photograph will be infinite even if this photograph is taken with a tiny telescope and with a very brief exposure. On the other hand, if at least two photons reaching the photographic plate are necessary to produce an image, then, no matter how large the telescope and no matter how long the exposure may be, the number of images of galaxies will be finite, with probability one.

REFERENCES

NEYMAN, J. (1955) : Sur la theorie probabiliste des amas de galaxies. *Ann. Inst. H. Poincare*, **14**, 201–244.

NEYMAN, J. and SCOTT, E. L. (1952) : A theory of the spatial distribution of galaxies. *Astrophys. J.*, **116**, 144–163.

Paper received : December, 1962.

EFFECTIVE ENTROPY RATE AND TRANSMISSION OF INFORMATION THROUGH CHANNELS WITH ADDITIVE RANDOM NOISE

By K. R. PARTHASARATHY

Indian Statistical Institute

SUMMARY. Transmission of information through channels with additive noise is considered. Coding theorem and its converse are established for these channels with a certain notion of capacity. This capacity is explicitly computed for this class of channels.

1. Introduction

The famous McMillan's theorem regarding ergodic sources can be reformulated as follows. Consider the minimum number of n-length sequences which have a total probability exceeding $1-\epsilon$. If μ is the measure describing the source denote this minimum by $N_n(\epsilon, \mu)$. McMillan's theorem states that for every ergodic source μ the limit $\lim\limits_{n\to\infty} \frac{\log N_n(\epsilon, \mu)}{n}$ exists and is always equal to the entropy rate of the source as defined by Shannon. The question arises as to what happens to the sequence $\frac{\log N_n(\epsilon, \mu)}{n}$ as $n\to\infty$ when the source is not necessarily ergodic but stationary. We show that except for a countable number of ϵ's the limit always exists and in general depends on ϵ. We construct two functions $A(\epsilon)$ and $B(\epsilon)$, $(0 < \epsilon < 1)$ which coincide except on a countable set and both $\underline{\lim}$ and $\overline{\lim}$ of $\frac{\log N_n(\epsilon, \mu)}{n}$ lie between $A(\epsilon)$ and $B(\epsilon)$. Further, as $\epsilon\to 0$, both the functions $A(\epsilon)$ and $B(\epsilon)$ converge to a unique limit $\overline{H}(\mu)$. The precise description of the functional $\overline{H}(\mu)$ is also given.

In the last section we introduce the notion of a channel with additive noise. Here the input and output alphabets coincide with a finite abelian group A and the noise is distributed according to an arbitrary stationary measure on the product space A'. When a message sequence is sent through the channel the noise gets added to the message independently of the message. The disturbed message is received at the output. The binary symmetric channel is a typical example. For the channel with additive noise distributed according to a stationary measure μ, we consider $M_n(\epsilon, \mu)$, the supremum of the length of all possible codes with probability of error less than or equal to ϵ (for transmission of messages during the time period $1, 2, \ldots, n$). A code of length N and probability of error less than or equal to ϵ is defined in the sense of Wolfowitz (1961). Then we analyse the asymptotic behaviour of the sequence $\frac{\log M_n(\epsilon, \mu)}{n}$. We show that the limit of this sequence exists for all ϵ except on a countable set. We also show that the $\underline{\lim}$ and $\overline{\lim}$ of this sequence lie between $\log a - A(\epsilon)$ and $\log a - B(\epsilon)$, where $A(\epsilon)$ and $B(\epsilon)$ are the functions mentioned in the previous paragraph and a is the number of elements in the alphabet A. As $\epsilon\to 0$

$\log a - A(\epsilon)$ and $\log a - B(\epsilon)$ converge to the same limit $\log a - \overline{H}(\mu)$. Thus in this case the capacity of the channel is not described by a single number but by the two functions $A(\epsilon)$ and $B(\epsilon)$.

The idea of studying the asymptotic properties of the sequence $\frac{\log N_n(\epsilon, \mu)}{n}$ is due to Winkelbauer. He gave the description of the function

$$\overline{H}(\mu) = \lim_{\epsilon \to 0} \overline{\lim_{n \to \infty}} \frac{\log N_n(\epsilon, \mu)}{n}$$

in terms of the entropies of the ergodic components of μ. It was stated by him without proof in his lecture at the Indian Statistical Institute.

2. Preliminaries

Throughout this paper A will denote a finite alphabet, A' the space of sequences of elements from A, T the shift transformation in A' and μ a measure which is defined on the usual σ-field of A' and invariant under T. We denote by $[x_1, x_2, \ldots, x_n]$ the cylinder set in A' of all sequences whose i-th coordinate is x_i for $i = 1, 2, \ldots, n$. Any n-length sequence $x_1, x_2, \ldots, x_n$ is referred to as a u-sequence. We denote by $N_n(\epsilon, \mu)$ the smallest number of u-sequences whose total probability is greater than or equal to $1-\epsilon$. This smallest set may not be unique. We choose one of them arbitrarily and denote it by $A_n(\epsilon, \mu)$.

If we assign the discrete topology to A and the product topology to A' then A' becomes a compact metric space. We shall now follow the notation of Oxtoby (1952). If $f(p)$ is a real valued function on A', let

$$M(f, p, k) = f_k(p) = \frac{1}{k} \sum_{i=1}^{k} f(T^i p) \qquad (k = 1, 2, \ldots)$$

and

$$M(f, p) = f^*(p) = \lim_{k \to \infty} M(f, p, k)$$

in case this limit exists. A Borel subset E of A' is said to have invariant measure one if $\mu(E) = 1$ for every invariant probability measure μ. Let Q be the set of points p for which $M(f, p)$ exists for every $f \in C(A')$ where $C(A')$ is the space of continuous functions on A'. It follows easily from Riesz's representation theorem that corresponding to any point $p \in Q$ there exists a unique invariant probability measure μ_p such that

$$M(f, p) = \int f \, d\mu_p.$$

Let $R \subset Q$ be the set of those points for which μ_p is ergodic. R is called the set of regular points. Then we have the following representation theorem of Kryloff and Bogoliouboff which can be found in Oxtoby (1952).

Theorem 2.1 : *The set R of regular points is a Borel measurable set of invariant measure one. For any Borel set $E \subset A'$, $\mu_p(E)$ is Borel measurable on R and*

$$\mu(E) = \int_R \mu_p(E) \, d\mu(p)$$

for any invariant probability measure μ.

Let $H(\mu)$ denote the entropy of any invariant probability measure μ.

Let

$$\overline{H}(\mu) = \text{ess sup} \quad H(\mu_p) \qquad \ldots \quad (2.1)$$

$$\underline{H}(\mu) = \text{ess inf} \quad H(\mu_p) \qquad \ldots \quad (2.2)$$

where the essential supremum and the essential infimum are taken relative to μ and μ_p denotes the ergodic measure corresponding to the regular point p.

3. Asymptotic properties of the function $N_n(\epsilon, \mu)$

In this section we shall prove the following theorem and two corollaries.

Theorem 3.1 : *Let $[A', \mu]$ be an arbitrary stationary source,*

then

$$A(\epsilon) \leqslant \underline{\lim} \frac{\log N_n(\epsilon, \mu)}{n} \leqslant \overline{\lim} \frac{\log N_n(\epsilon, \mu)}{n} \leqslant B(\epsilon) \qquad \ldots \quad (3.1)$$

where

$$A(\epsilon) = \lim_{\downarrow} \eta(\delta), \qquad \ldots \quad (3.2)$$

$$B(\epsilon) = \lim_{\delta \uparrow \epsilon} \eta'(\delta), \qquad \ldots \quad (3.3)$$

$\eta(\delta)$ is the greatest number with the property

$$\mu[p : H(\mu_p) \geqslant \eta] \geqslant \delta$$

and $\eta'(\delta)$ is the smallest number with the property

$$\mu[p : H(\mu_p) \leqslant \eta'] \geqslant 1-\delta.$$

Corollary 3.1 : *For any stationary source $[A', \mu]$, $\lim\limits_{n \to \infty} \frac{\log N_n(\epsilon, \mu)}{n}$ exists for all $0 < \epsilon < 1$ except for a countable set.*

Corollary 3.2 : *For any stationary source $[A', \mu]$,*

$$\lim_{\varepsilon \to 0} \underline{\lim}_{n \to \infty} \frac{\log N_n(\epsilon, \mu)}{n} = \lim_{\varepsilon \to 0} \overline{\lim}_{n \to \infty} \frac{\log N_n(\epsilon, \mu)}{n} = \overline{H}(\mu).$$

Before proceeding to the proof of Theorem 3.1 we need to establish two lemmas.

Lemma 3.1 : *For any stationary source $[A', \mu]$, the limit*

$$\lim_{n \to \infty} -\frac{1}{n} \log \mu[x_1 \ldots x_n] = g_\mu(x)$$

exists in measure and $\qquad g_\mu(p) = H(\mu_p) \quad$ a.e. $p(\mu)$.

Proof: The existence of the limit is the famous McMillan's theorem. In the course of Khinchin's proof of McMillan's theorem (1957) it can be seen that $g_\mu(x)$ can be obtained as follows. Define $h_\mu(x)$ as the conditional probability of x_0 given $x_{-1}, x_{-2}, \ldots$ under μ. (Here it is assumed that $x = (\ldots x_{-1}, x_0, x_1, \ldots)$).

Then

$$g_\mu(x) = \lim_{n\to\infty} \frac{h_\mu(x)+h_\mu(Tx)+\ldots+h_\mu(T^{n-1}x)}{n} \quad \text{a.e.} \ (\mu).$$

But by Theorem 2.6 of the author (1961),

we have

$$h_\mu(x) = h_{\mu_p}(x) \quad \text{a.e.} \quad x(\mu_p)$$

for almost all $p(\mu)$.

Further

$$\lim_{n\to\infty} \frac{h_{\mu_p}(x)+h_{\mu_p}(Tx)+\ldots+h_{\mu_p}(T^{n-1}x)}{n} = H(\mu_p) \ \text{a.e.} \ (\mu_p)$$

for any regular point p.

Thus

$$g_\mu(x) = H(\mu_p) \quad \text{a.e.} \quad x(\mu_p)$$

for almost all $p(\mu)$. From the Kryloff-Bogoliouboff theory of regular points in a dynamical system we have

$$\mu_x = \mu_p \qquad \text{a.e.} \quad x(\mu_p).$$

Thus

$$g_\mu(x) = H(\mu_x) \quad \text{a.e.} \quad (\mu_p)$$

for almost all $p(\mu)$. Thus the set $E = [x : g_\mu(x) \neq H(\mu_x)]$ has measure zero under μ_p for almost all $p(\mu)$. An application of Theorem 2.1 shows that

$$\mu(E) = \int \mu_p(E) d\mu(p) = 0.$$

This completes the proof of Lemma 3.1.

Lemma 3.2 : *For any stationary source* $[A', \mu]$ *and any* $\epsilon > 0$

$$\underline{H}(\mu) \leqslant \varliminf_{n\to\infty} \frac{\log N_n(\epsilon, \mu)}{n} \leqslant \varlimsup_{n\to\infty} \frac{\log N_n(\epsilon, \mu)}{n} \leqslant \overline{H}(\mu).$$

Proof : From Lemma 3.1 it is clear that

$$\underline{H}(\mu) \leqslant g_\mu(x) \leqslant \overline{H}(\mu)$$

with probability one. Thus if we write, for any fixed $\eta > 0$,

$$\mu[x : -\frac{1}{n} \log \mu(x_1 \ldots x_n) \geqslant \underline{H}(\mu) - \eta] = 1 - \delta_n \qquad \ldots \quad (3.4)$$

then $\delta_n \to 0$ as $n \to \infty$. The complement of the set written within braces in (3.4) has probability δ_n. Any set with probability $> 1-\epsilon$ must have a subset with pro-

bability greater than $1-\epsilon-\delta_n$ whose elements satisfy the inequality within braces in (3.4). Suppose this subset has N' u-sequences. For these sequences

$$\mu(x_1, x_2, \ldots, x_n) \leqslant 2^{-n(\underline{H}(\mu)-\eta)}.$$

Summing up over this subset, we get

$$1-\epsilon-\delta_n \leqslant N' 2^{-n(\underline{H}(\mu)-\eta)}.$$

Thus
$$1-\epsilon-\delta_n \leqslant N_n(\epsilon, \mu) 2^{-n(\underline{H}(\mu)-\eta)}.$$

Since $\delta_n \to 0$ as $n \to \infty$ after some stage

$$1-\epsilon-\delta_n \geqslant \frac{1-\epsilon}{2}.$$

Thus
$$\frac{\log N_n(\epsilon, \mu)}{n} \geqslant \frac{\log (1-\epsilon)/2}{n} + \underline{H}(\mu) - \eta$$

which implies
$$\varliminf \frac{\log N_n(\epsilon, \mu)}{n} \geqslant \underline{H}(\mu) - \eta.$$

Since η is arbitrary we have

$$\varliminf \frac{\log N_n(\epsilon, \mu)}{n} \geqslant \underline{H}(\mu).$$

In order to prove the other inequality, consider, for any fixed $\eta > 0$, the sequence of numbers

$$\mu[x : -\frac{1}{n} \log \mu(x_1 \ldots x_n) \leqslant \underline{H}(\mu) + \eta] = 1 - \delta_n'. \qquad \ldots (3.5)$$

By Lemma 3.1 and (2.1) we have $\lim_{n \to \infty} \delta_n' = 0.$

Thus there exists a subset A_n of u-sequences satisfying the inequality within braces in (3.5) whose probability exceeds $1-\epsilon$ for all sufficiently large n. If the inequality within braces in (3.5) is satisfied, then

$$\mu[x_1 \ldots x_n] > 2^{-n(\bar{H}(\mu)+\eta} . \qquad \ldots (3.6)$$

If N' u-sequences satisfying (3.6) are required to make up a probability greater than $1-\epsilon$ then

$$N_n(\epsilon, \mu) \leqslant N' \leqslant 2^{n(\bar{H}(\mu)+\eta)} .$$

Thus
$$\varlimsup \frac{\log N_n(\epsilon, \mu)}{n} \leqslant \bar{H}(\mu) + \eta.$$

The arbitrariness of η implies the validity of the lemma.

Then

$$g_\mu(x) = \lim_{n\to\infty} \frac{h_\mu(x)+h_\mu(Tx)+\ldots+h_\mu(T^{n-1}x)}{n} \text{ a.e. } (\mu).$$

But by Theorem 2.6 of the author (1961),

we have
$$h_\mu(x) = h_{\mu_p}(x) \quad \text{a.e.} \quad x(\mu_p)$$

for almost all $p(\mu)$.

Further
$$\lim_{n\to\infty} \frac{h_{\mu_p}(x)+h_{\mu_p}(Tx)+\ldots+h_{\mu_p}(T^{n-1}x)}{n} = H(\mu_p) \text{ a.e. } (\mu_p)$$

for any regular point p.

Thus
$$g_\mu(x) = H(\mu_p) \quad \text{a.e.} \quad x(\mu_p)$$

for almost all $p(\mu)$. From the Kryloff-Bogoliouboff theory of regular points in a dynamical system we have

$$\mu_x = \mu_p \qquad \text{a.e.} \quad x(\mu_p).$$

Thus
$$g_\mu(x) = H(\mu_x) \quad \text{a.e.} \quad (\mu_p)$$

for almost all $p(\mu)$. Thus the set $E = [x : g_\mu(x) \neq H(\mu_x)]$ has measure zero under μ_p for almost all $p(\mu)$. An application of Theorem 2.1 shows that

$$\mu(E) = \int \mu_p(E)d\mu(p) = 0.$$

This completes the proof of Lemma 3.1.

Lemma 3.2 : *For any stationary source $[A', \mu]$ and any $\epsilon > 0$*

$$\underline{H}(\mu) \leqslant \varliminf_{n\to\infty} \frac{\log N_n(\epsilon, \mu)}{n} \leqslant \varlimsup_{n\to\infty} \frac{\log N_n(\epsilon, \mu)}{n} \leqslant \overline{H}(\mu).$$

Proof: From Lemma 3.1 it is clear that

$$\underline{H}(\mu) \leqslant g_\mu(x) \leqslant \overline{H}(\mu)$$

with probability one. Thus if we write, for any fixed $\eta > 0$,

$$\mu[x : -\frac{1}{n} \log \mu(x_1 \ldots x_n) \geqslant \underline{H}(\mu)-\eta] = 1-\delta_n \qquad \ldots \quad (3.4)$$

then $\delta_n \to 0$ as $n\to\infty$. The complement of the set written within braces in (3.4) has probability δ_n. Any set with probability $> 1-\epsilon$ must have a subset with pro-

bability greater than $1-\epsilon-\delta_n$ whose elements satisfy the inequality within braces in (3.4). Suppose this subset has N' u-sequences. For these sequences

$$\mu[x_1, x_2, \ldots, x_n] \leqslant 2^{-n(\underline{H}(\mu)-\eta)}.$$

Summing up over this subset, we get

$$1-\epsilon-\delta_n \leqslant N' 2^{-n(\underline{H}(\mu)-\eta)}.$$

Thus
$$1-\epsilon-\delta_n \leqslant N_n(\epsilon, \mu) 2^{-n(\underline{H}(\mu)-\eta)}.$$

Since $\delta_n \to 0$ as $n \to \infty$ after some stage

$$1-\epsilon-\delta_n \geqslant \frac{1-\epsilon}{2}.$$

Thus
$$\frac{\log N_n(\epsilon, \mu)}{n} \geqslant \frac{\log (1-\epsilon)/2}{n} + \underline{H}(\mu) - \eta$$

which implies
$$\varliminf \frac{\log N_n(\epsilon, \mu)}{n} \geqslant \underline{H}(\mu) - \eta.$$

Since η is arbitrary we have

$$\varliminf \frac{\log N_n(\epsilon, \mu)}{n} \geqslant \underline{H}(\mu).$$

In order to prove the other inequality, consider, for any fixed $\eta > 0$, the sequence of numbers

$$\mu[x : -\frac{1}{n} \log \mu(x_1 \ldots x_n) \leqslant \underline{H}(\mu) + \eta] = 1 - \delta_n'. \qquad \ldots \quad (3.5)$$

By Lemma 3.1 and (2.1) we have $\lim_{n \to \infty} \delta_n' = 0.$

Thus there exists a subset A_n of u-sequences satisfying the inequality within braces in (3.5) whose probability exceeds $1-\epsilon$ for all sufficiently large n. If the inequality within braces in (3.5) is satisfied, then

$$\mu[x_1 \ldots x_n] > 2^{-n(\bar{H}(\mu)+\eta} . \qquad \ldots \quad (3.6)$$

If N' u-sequences satisfying (3.6) are required to make up a probability greater than $1-\epsilon$ then

$$N_n(\epsilon, \mu) \leqslant N' \leqslant 2^{n(\bar{H}(\mu)+\eta)} .$$

Thus
$$\varlimsup \frac{\log N_n(\epsilon, \mu)}{n} \leqslant \bar{H}(\mu) + \eta.$$

The arbitrariness of η implies the validity of the lemma.

We shall now turn to the proof of Theorem 3.1. Choose any $\delta > \epsilon$. Choose the largest η such that

$$\mu[p : H(\mu_p) \geqslant \eta] \geqslant \delta.$$

Let it be $\eta(\delta)$.

If
$$E = [p : H(\mu_p) \geqslant \eta(\delta)]$$
then
$$\mu(E) \geqslant \delta.$$

Define
$$\mu_1(B) = \frac{\mu(B \cap E)}{\mu(E)}, \qquad \mu_2(B) = \frac{\mu(B \cap E')}{\mu(E')}.$$

We assume that $\mu(E') > 0$.

Then
$$\mu = a\mu_1 + (1-a)\mu_2$$
where
$$a = \mu(E) \geqslant \delta > \epsilon.$$

If we consider the set $A_n(\epsilon, \mu)$ (the smallest set of u-sequences with probability $> 1-\epsilon$)

then
$$\mu_1(A_n(\epsilon, \mu)) \geqslant \frac{1-\epsilon-(1-a)}{a} \geqslant \frac{a-\epsilon}{a} \geqslant \frac{\delta-\epsilon}{a}.$$

Thus
$$N_n(\epsilon, \mu) \geqslant N_n\left(1-\frac{\delta-\epsilon}{a}, \mu_1\right)$$

and
$$0 < 1-\frac{\delta-\epsilon}{a} < 1.$$

An application of Lemma 3.1 shows that

$$\underline{\lim} \frac{\log N_n(\epsilon, \mu)}{n} \geqslant \underline{H}(\mu_1) \geqslant \eta(\delta).$$

If $\mu(E') = 0$ this inequality is trivially valid. Since δ is any number $> \epsilon$ and $\eta(\delta)$ increases to $A(\epsilon)$ as δ descends to ϵ

we have
$$\underline{\lim} \frac{\log N_n(\epsilon, \mu)}{n} \geqslant A(\epsilon). \qquad \ldots (3.7)$$

For proving the other inequality choose $\delta < \epsilon$ and then choose the smallest η' with the property

$$\mu[p : H(\mu_p) \leqslant \eta'] \geqslant 1-\delta.$$

Let it be $\eta'(\delta)$.

Let
$$F = [p : H(\mu_p) \leqslant \eta'(\delta)],$$

and
$$\mu_1'(B) = \frac{\mu(B \cap F)}{\mu(F)}, \qquad \mu_2'(B) = \frac{\mu(B \cap F')}{\mu(F')}.$$

Then
$$\mu = b\mu_1' + (1-b)\mu_2'$$

where $b = \mu(F) \geqslant 1-\delta$. If $\mu_1'(B) > \frac{1-\epsilon}{b}$ then $\mu(B) > 1-\epsilon$.

Thus
$$N_n(\epsilon, \mu) \leqslant N_n\left(1-\frac{1-\epsilon}{b}, \mu_1'\right),$$

$$0 < 1-\frac{1-\epsilon}{b} < 1.$$

Thus by Lemma 3.2
$$\overline{\lim}\ \frac{\log N_n(\epsilon, \mu)}{n} \leqslant \bar{H}(\mu_1') \leqslant \eta'(\delta).$$

As δ increases to ϵ, $\eta'(\delta)$ decreases to a limit $B(\epsilon)$.

Thus
$$\overline{\lim}\ \frac{\log N_n(\epsilon, \mu)}{n} \leqslant B(\epsilon). \qquad \ldots \quad (3.8)$$

(3.7) and (3.8) complete the proof of Theorem 3.1. Corollary 3.1 is an immediate consequence of the fact that in the real line there cannot be more than a countable collection of disjoint open intervals and hence $A(\epsilon) = B(\epsilon)$ except for a countable set. Corollary 3.2 follows immediately from the fact that $A(\epsilon)$ and $B(\epsilon)$ converge to $\bar{H}(\mu)$ as $\epsilon \to 0$.

Remark : From Theorem 3.1 and Corollary 3.2 it is clear that the number $\bar{H}(\mu)$ defined by (2.1) can be rightly called the effective entropy rate of the stationary source $[A', \mu]$. The result that

$$\lim_{\epsilon \to 0}\ \overline{\lim_{n\to\infty}}\ \frac{\log N_n(\epsilon, \mu)}{n} = \bar{H}(\mu)$$

is due to K. Winkelbauer. It was stated by him without proof in one of his lectures at the Indian Statistical Institute. It was his conjecture that

$$\lim_{n\to\infty} \frac{\log N_n(\epsilon, \mu)}{n} \text{ exists for every } \epsilon.$$

4. Channels with additive noise

In this section we introduce the notion of a stationary channel with additive noise, define its capacity and prove the coding theorem as well as its converse.

Let the input and output alphabets of a channel coincide with a finite abelian group A. In a natural way the space A' becomes an abelian group. We denote by $+$ and $-$ the addition and inverse operation in the group A'. For any set E we write

$$E-x = [z : z \epsilon A', z+x \epsilon E].$$

Let μ_0 be an invariant measure defined on A'. Then the probability distributions

$$\nu_x(F) = \mu(F-x) \qquad \ldots \quad (4.1)$$

where F is any set in the usual σ-field of A' and x is any point in A' define a stationary channel. The distribution at the output of this channel corresponding to any input distribution λ is obtained by convoluting λ with μ. Even if μ is ergodic this channel need not be of finite memory in the sense of Feinstein. If the group A consists of two elements 0 and 1 only, the addition is done modulo 2 and μ is the product measure obtained by assuming probability p for one, then we get the so-called binary symmetric channel. We shall call a channel whose distributions are specified by (4.1) as a channel with additive noise and noise distribution μ.

A code with error ϵ and length N is a collection of u-sequences $u_1, u_2, \ldots, u_N$ and sets $V_1, V_2, \ldots, V_N$ of u-sequences with the properties

$$\text{(i)} \quad \mu(V_i - u_i) > 1-\epsilon$$

$$\text{(ii)} \quad V_i \cap V_j = \phi \text{ for } i \neq j. \qquad \ldots \quad (4.2)$$

Let $M_n(\epsilon, \mu)$ be the maximal length possible for a code with error ϵ for a channel with additive noise and noise distribution μ. Then we have the following theorem.

Theorem 4.1 : *For a stationary channel with additive noise and noise distribution* μ

$$\log_2 a - B(\epsilon) \leqslant \underline{\lim} \frac{\log M_n(\epsilon, \mu)}{n} \leqslant \overline{\lim} \frac{\log M_n(\epsilon, \mu)}{n} \leqslant \log_2 a - A(\epsilon)$$

where $A(\epsilon)$ *and* $B(\epsilon)$ *are the functions occurring in the statement of Theorem* 3.1.

Corollary 4.1 : *Except for a countable set of* ϵ*'s the limit*

$$\lim_{n\to\infty} \frac{\log M_n(\epsilon, \mu)}{n} = \log_2 a - A(\epsilon) = \log_2 a - B(\epsilon)$$

exists. Further

$$\lim_{\epsilon\to 0} \log a - A(\epsilon) = \lim_{\epsilon \to 0} \log a - B(\epsilon) = \log a - \overline{H}(\mu).$$

Remark : Corollary 4.1 justifies our calling $\log a - \overline{H}(\mu)$ as the capacity of the additive channel with noise distribution μ.

Proof of Theorem 4.1 : Let $u_1, u_2, \ldots, u_N, V_1, \ldots, V_N$ be a code with error ϵ. For any set E of u-sequences let $m(E)$ be the number of u-sequences in E. By property (i) of (4.2) we have

$$m(V_i) \geqslant N_n(\epsilon, \mu). \qquad \ldots \quad (4.3)$$

For any $\delta > 0$ and all sufficiently large n, we have by Theorem 3.1 and (4.3)

$$m(V_i) \geqslant 2^{n[A(\epsilon)-\delta]}.$$

Since V_i's are disjoint and the total number of u-sequences is a^n, we obtain

$$a^n \geqslant m(\bigcup_1^N V_i) \geqslant N \, . \, 2^{n[A(\epsilon)-\delta]}.$$

Thus
$$\frac{\log N}{n} \leqslant \log a - A(\epsilon) + \delta. \quad \ldots \quad (4.4)$$

Since (4.4) is true for any code of error ϵ, we get

$$\frac{\log M_n(\epsilon, \mu)}{n} \leqslant \log a - A(\epsilon) + \delta.$$

Allowing n to tend to infinity and then noting the arbitrariness of δ, we get

$$\overline{\lim} \frac{\log M_n(\epsilon, \mu)}{n} \leqslant \log a - A(\epsilon).$$

For proving the other inequality in the theorem we follow Takano (1957). Let $\epsilon' > 0$ be an arbitrary number less than ϵ. Choose the set V_1 to be the set with the smallest number of u-sequences whose probability exceeds $1-\epsilon'$. In the notation given in Section 1, $V_1 = A_n(\epsilon', \mu)$. Let u_1 be the u-sequence all of whose elements coincide with the identity of the group A. Now choose u_2 such that

$$\mu[(V_1+u_2)V_1'-u_2] > 1-\epsilon$$

where the prime is used to denote the complement. If no such u_2 exists stop. Write

$$V_2 = (V_1+u_2)V_1'.$$

Choose u_3 such that

$$\mu[(V_1+u_3)V_2'V_1'-u_3] > 1-\epsilon.$$

If no such u_3 exists stop. At the r-th stage u_r is chosen such that

$$\mu[(V_1+u_r)V_{r-1}'V_{r-2}' \ldots V_1'-u_r] > 1-\epsilon.$$

Then we write
$$V_r = (V_1+u_r)V_{r-1}' V_{r-2}' \ldots V_1'.$$

Let the process terminate after N stages. Let $V = \bigcup_i V_i$. Then we have u-sequences $u_1, u_2, \ldots u_N$ and sets $V_1, V_2, \ldots, V_N$ of u-sequences with the properties

(1) $\mu(V_i-u_i) > 1-\epsilon$,

(2) $V_i \subseteq V_1+u_i$,

(3) $V_i \cap V_j = \phi$,

(4) $V = \bigcup_i V_i = \bigcup_i (V_1+u_i)$,

(5) For any u-sequence $\mu((V_1+u)V'-u) \leqslant 1-\epsilon$.

Since V_1 has probability greater than $1-\epsilon'$, we have from property (5) above,

$$\begin{aligned} 1-\epsilon' \leqslant \mu(V_1) &= \mu(V_1+u-u) \\ &\leqslant \mu((V_1+u)V-u)+\mu((V_1+u)V'-u) \\ &\leqslant \mu(V-u)+(1-\epsilon). \end{aligned}$$

This inequality can be rewritten as

$$\mu(V-u) \geqslant \epsilon-\epsilon'. \qquad \dots \quad (4.5)$$

Since (4.5) is true for every u, we obtain by multiplying both sides of (4.5) by a^{-n} and adding over all the u-sequences

$$m(V)\, a^{-n} \geqslant \epsilon-\epsilon'$$

or

$$m(V) \geqslant a^n . (\epsilon-\epsilon'). \qquad \dots \quad (4.6)$$

Since

$$V = \bigcup_1^N V_i \text{ and } V_i \sqsubset V_1 + u_i,$$

we have

$$m(V) \leqslant N . m\,(V_1). \qquad \dots \quad (4.7)$$

We recall that $V_1 = A_n(\epsilon', \mu)$. By an application of Theorem 3.1 we have, for any $\delta > 0$ and all sufficiently large n

$$m(V_1) \leqslant 2^{n[B(\epsilon')+\delta]}. \qquad \dots \quad (4.8)$$

Combining (4.6), (4.7) and (4.8)

$$N \geqslant (\epsilon-\epsilon') \cdot a^n \cdot 2^{-n[B(\epsilon')+\delta]}.$$

Since $M_n(\epsilon, \mu) \geqslant N$ we have

$$\frac{\log M_n(\epsilon, \mu)}{n} \geqslant \frac{\log (\epsilon-\epsilon')}{n} + [\log a - B(\epsilon')] - \delta.$$

Allowing $n \to \infty$ and then $\delta \to 0$, we get

$$\varliminf_{n\to\infty} \frac{\log M_n(\epsilon, \mu)}{n} \geqslant \log a - B(\epsilon').$$

From the definition of the function $B(\epsilon)$ we see that it is left continuous. Since ϵ' is any number less than ϵ, we get by letting ϵ' increase to ϵ

$$\varliminf_{n\to\infty} \frac{\log M_n(\epsilon, \mu)}{n} \geqslant \log a - B(\epsilon).$$

Corollary 4.1 is an immediate consequence of Corollaries 3.1 and 3.2.

REFERENCES

KHINCHIN, A. I. (1957): *Mathematical Foundations of Information Theory*, Dover Publications, New York.

OXTOBY, J. C. (1952): Ergodic sets. *Bull. Amer. Math. Soc.*, **58**, 116-136.

PARTHASARATHY, K. R. (1961): On the integral representation of the rate of transmission of a stationary channel. *Illinois J. Math.*, **5**, 299-305.

TAKANO, K. (1957): On the basic theorems of information theory. *Ann. Inst. Stat. Math.*, **9**, 53-77.

WOLFOWITZ, J. (1961): *Coding Theorems of Information Theory*, Springer-Verlag, Berlin.

Paper received : September, 1962.

A STABILITY THEOREM FOR THE BINOMIAL LAW*

By B. RAMACHANDRAN

Institute of Mathematical Sciences, Madras

SUMMARY. It is shown that if the sum of two independent (one-dimensional) random variables is "approximately" Binomial, the same is true of the individual summands also, the degree of closeness of two random variables being measured by the distance function

$$d(F, G) = \sup_x |F(x) - G(x)|$$

between their respective distribution functions F and G. Analogous results are already known for the Normal and the Poisson laws.

1. Introduction

Let X_1 and X_2 be independent (real-valued) random variables, and let $X = X_1 + X_2$.

In the theory of decomposition of random variables, it is well known that if X has a Normal (respectively a Poisson, a Binomial) distribution, then the same is true of the summands X_1 and X_2 also. Details on these results which are due respectively to P. Lévy -H. Cramér, D. A. Raikov, and N. A. Sapogov-H. Teicher can be found, for instance, in Lukacs (1960), Chapter 8.

Sapogov (1951 and 1959) established a 'stability theorem' for the Normal law, namely, that if X has a distribution which is sufficiently close to a Normal distribution, then the same is true of X_1 and X_2 also, the degree of closeness between two distribution functions being measured by the 'Kolmogorov distance'

$$d(F, G) = \sup_x |F(x) - G(x)|.$$

An analogous result for the Poisson law was established by Shalaevsky (1959). It is the purpose of this paper to establish a similar result for the Binomial law.

2. Some notations

We shall use the following notations.

For all real x,

$$B(x;\ n, p) = \sum_{r \leqslant x} \binom{n}{r} p^r q^{n-r}$$

r running through all integer-values, where we define (for convenience of notation):

$$\binom{n}{r} = 0 \quad \text{for} \quad r < 0 \text{ and } r > n.$$

$F(x)$, $F_1(x)$ and $F_2(x)$ will respectively denote the distribution functions of X, X_1 and X_2. (All distribution functions will be assumed to be right-continuous).

* Research supported by the Council of Scientific and Industrial Research, Government of India, under their, "Scientific Pool" scheme.

3. Statement and proof of the theorem

Theorem : *Let $\epsilon > 0$ be a sufficiently small number (the degree of smallness required will be made clear in course of the proof), and let*

$$\sup_x |F(x)-B(x; n, p)| < \epsilon. \qquad \dots (3.1)$$

Then there exist integers n_1, n_2, and a positive constant k independent of ϵ, such that

$$0 \leqslant n_1,\ n_2 \leqslant n;\ n_1+n_2 = n \qquad \dots (3.2)$$

$$\sup_x |F_1(x)-B(x-c; n_1, p)| < k\epsilon^{1/2n}, \qquad \dots (3.3)$$

$$\sup_x |F_2(x)-B(x+c; n_2, p)| < k\epsilon^{1/2n} \qquad \dots (3.4)$$

where

$$c = \sup \{x | P(X_1 < x) < \sqrt{\epsilon}\}. \qquad \dots (3.5)$$

Proof: We shall assume without loss of generality that $c = 0$. (Otherwise, we need only consider the random variables X_1-c and X_2+c instead of X_1 and X_2 respectively.)

Then we have

$$P(X_1 < 0) \leqslant \sqrt{\epsilon}. \qquad \dots (3.6)$$

Again, for any $\delta > 0$,

$$P(X_1 < \delta) \geqslant \sqrt{\epsilon}.$$

Hence,

$$P(X_1 \leqslant 0) \geqslant \sqrt{\epsilon}. \qquad \dots (3.7)$$

From (3.1), we have in particular that

$$P(X < 0) < \epsilon.$$

Hence

$$\sqrt{\epsilon}\, P(X_2 < 0) \leqslant P(X_1 \leqslant 0)\, P(X_2 < 0) \leqslant P(X < 0) < \epsilon,$$

so that

$$P(X_2 < 0) < \sqrt{\epsilon}. \qquad \dots (3.8)$$

Now, for $j, k = 1, 2;\ j \neq k$, the event $(X_j > 0,\ X_k \geqslant 0)$ implies the event $(X > 0)$. Hence

$$(X \leqslant 0) \subset (X_j \leqslant 0) \cup (X_k < 0),$$

so that $P(X \leqslant 0) \leqslant P(X_j \leqslant 0)+P(X_k < 0) = P(X_1 < 0)+P(X_2 < 0)+P(X_j = 0)$ for $j = 1, 2$. Hence, from (3.1), (3.6) and (3.8), we have for $j = 1, 2$,

$$q^n-\epsilon < P(X \leqslant 0) < 2\sqrt{\epsilon}+P(X_j = 0)$$

which implies that

$$P(X_j = 0) > \tfrac{1}{2}\, q^n \quad (j = 1, 2) \qquad \dots (3.9)$$

ϵ being assumed to be so small that

$$\epsilon+2\sqrt{\epsilon} < \tfrac{1}{2}q^n. \qquad \dots (3.10)$$

Hence, for

$$j, k = 1, 2,; \quad j \neq k,$$

we have

$$\tfrac{1}{2}q^n\, P(X_j < 0) < P(X_k = 0)\, P(X_j < 0) \leqslant P(X < 0) < \epsilon$$

so that

$$P(X_j < 0) < 2\epsilon q^{-n} \quad (j = 1, 2). \qquad \dots (3.11)$$

Remark : However, the estimates (3.11) for $P(X_j < 0)$ are more of interest than of importance to us : the estimates (3.6) and (3.8) are quite adequate for our purposes as will be seen during the derivation of relation (3.25).

Now, from (3.1), it follows that for all real x,

$$|F(x-0)-B(x-0; n, p)| = \lim_{y \to x-0} |F(y)-B(y; n, p)| \leqslant \epsilon.$$

Hence, for all integers r.

$$\begin{aligned} & |P(r \leqslant X < r+1) - \binom{n}{r} p^r q^{n-r}| \\ & = |[F(r+1-0)-F(r-0)]-[B(r+1-0; n,p)-B(r-0; n, p)]| \\ & \leqslant |F(r+1-0)-B(r+1-0; n, p)| + |F(r-0)-B(r-0; n, p)| \leqslant 2\epsilon. \quad \ldots \quad (3.12) \end{aligned}$$

By a similar reasoning,

$$P(r < X < r+1) < 2\epsilon. \qquad \ldots \quad (3.13)$$

Hence, for $j, k = 1, 2; j \neq k$, and for all integers r, by (3.9),

$$\begin{aligned} \tfrac{1}{2}q^n P(r < X_j < r+1) &< P(X_k = 0)\, P(r < X_j < r+1) \\ &\leqslant P(r < X < r+1) < 2\epsilon \end{aligned}$$

whence

$$P(r < X_j < r+1) < 4\epsilon q^{-n} \qquad (j = 1, 2). \qquad \ldots \quad (3.14)$$

We now define

$$a = \inf \{x | P(X_1 > x) < 2\sqrt{\epsilon}\, q^{-n}\}. \qquad \ldots \quad (3.15)$$

We will assume ϵ to be so small that $a \geqslant c$ (= 0 by our assumption).

From the definition of a,

$$P(X_1 > a) \leqslant 2\sqrt{\epsilon}\, q^{-n} \qquad \ldots \quad (3.16)$$

and

$$P(X_1 > x) \geqslant 2\sqrt{\epsilon}\, q^{-n} \quad \text{for all } x < a$$

whence

$$P(X_1 \geqslant a) \geqslant 2\sqrt{\epsilon}\, q^{-n}. \qquad \ldots \quad (3.17)$$

We note that, since $P(X > n) < \epsilon$ from (3.1), and $P(X_2 = 0) > \frac{1}{2}q^n$ from (3.9),

$$P(X_1 > n) \leqslant P(X > n)/P(X_2 = 0) < 2\epsilon\, q^{-n} < 2\sqrt{\epsilon}\, q^{-n}$$

so that

$$n \geqslant a. \qquad \ldots \quad (3.18)$$

Let

$$\left.\begin{aligned} n_1 = [a], \quad \text{and} \quad n_2 = n - n_1 \\ 0 \leqslant n_1, n_2 \leqslant n; \quad n_1 + n_2 = n \end{aligned}\right\} \qquad \ldots \quad (3.19)$$

so that

(As usual, $[x]$ denotes the largest integer not greater than x.)

We note the further facts below :

$$P(n_1 < X_1 < n_1+1) < 4\epsilon q^{-n};$$

by (3.14),

and, since $a < n_1 + 1$, we have by (3.16),

$$P(X_1 \geqslant n_1+1) \leqslant P(X_1 > a) \leqslant 2\sqrt{\epsilon}\, q^{-n}.$$

Hence

$$P(X_1 > n_1) < 6\sqrt{\epsilon}\, q^{-n}. \qquad \ldots \quad (3.20)$$

Also, since $n_1 \leqslant a$,

$$P(X_1 \geqslant n_1) \geqslant P(X_1 \geqslant a) \geqslant 2\sqrt{\epsilon}\, q^{-n} \text{ from (3.17),}$$

whence $$P(X_2 > n_2) \leqslant P(X > n)/P(X_1 \geqslant n_1) < \tfrac{1}{2}\sqrt{\epsilon}\, q^n. \quad \ldots \quad (3.21)$$

From (3.20) and (3.21), it follows that ($c_1, c_2, \ldots,$ denoting positive constants not depending on ϵ) for $j = 1, 2$

$$P(X_j > n_j) < c_1 \sqrt{\epsilon}. \quad \ldots \quad (3.22)$$

We now define the new random variables (for $j = 1, 2$)

$$Y_j = \begin{cases} 0 & \text{if } X_j < 0 \text{ or } > n_j \\ r & \text{if } r \leqslant X_j < r+1, \text{ for } 0 \leqslant r < n_j \\ n_j & \text{if } X_j = n_j. \end{cases}$$

Then, the independence of X_1 and X_2 implies that of Y_1 and Y_2. Let $Y = Y_1 + Y_2$.

We estimate the quantity

$$\begin{aligned} D_r &= P(r \leqslant X < r+1) - P(r \leqslant Y < r+1) \\ &= P(r \leqslant X < r+1) - P(Y = r) \quad \ldots \quad (3.23) \end{aligned}$$

where r is an arbitrary integer.

Let $$P_r = P(r \leqslant X < r+1;\ X \neq Y)$$

and $$P'_r = P(Y = r;\ X \neq Y).$$

It is clear that $$D_r = P_r - P'_r; \text{ and } P_r, P'_r \leqslant P(X \neq Y)$$

and hence that $$|D_r| \leqslant P_r + P'_r \leqslant 2P(X \neq Y). \quad \ldots \quad (3.24)$$

We therefore proceed to estimate $P(X \neq Y)$: since the event $(X_1 = Y_1;\ X_2 = Y_2)$ implies the event $(X = Y)$, it follows that

$$(X \neq Y) \subset (X_1 \neq Y_1) \cup (X_2 \neq Y_2)$$

so that $$\begin{aligned} P(X \neq Y) &\leqslant P(X_1 \neq Y_1) + P(X_2 \neq Y_2) \\ &= \sum_{j=1}^{2} \Big[P(X_j < 0) + P(X_j > n_j) + \sum_{r=0}^{n_j - 1} P(r < X_j < r+1)\Big] \\ &< \sum_{j=1}^{2} (2\epsilon q^{-n} + c_1\sqrt{\epsilon} + 4\epsilon q^{-n}.\, n_j), \end{aligned}$$

using (3.11), (3.22) and (3.14) respectively,

$$< c_2\sqrt{\epsilon}. \quad \ldots \quad (3.25)$$

Hence, from (3.24) and (3.25),

$$|D_r| < 2c_2\sqrt{\epsilon}. \quad \ldots \quad (3.26)$$

We then have from (3.12) and (3.26) that for all integers r (recalling the notation in Section 2)

$$\left|P(Y = r) - \binom{n}{r} p^r q^{n-r}\right| < c_3\sqrt{\epsilon}. \quad \ldots \quad (3.27)$$

Let now $g(z)$, $g_1(z)$ and $g_2(z)$ be the probability-generating functions of the variables Y, Y_1 and Y_2 respectively. They are polynomials of degrees n, n_1 and n_2 respectively, and (for all complex z)

$$g_1(z) \cdot g_2(z) = g(z) = (q+pz)^n + \sum_{r=0}^{n} \epsilon_r z^r \qquad \dots \quad (3.28)$$

where, by (3.27),

$$|\epsilon_r| < c_3\sqrt{\epsilon} \,. \qquad \dots \quad (3.29)$$

We present the argument henceforward in rather condensed form, to avoid trivial and at the same time cumbersome details.

Let these three polynomials in z transform respectively into the polynomials $f(w)$, $f_1(w)$ and $f_2(w)$, under the transformation $w = z+(q/p)$. By a straight forward application of Rouche's theorem on the zeros of analytic functions (see, for instance, Titchmarsh (1939, p. 116)), it is easily seen that (if ϵ is sufficiently small) all the zeros of $f(w) \equiv g(z)$ lie in the circle $|w| < [c_4\sqrt{\epsilon}/(p^n+\epsilon_n)]^{1/n}$, so that, if ϵ be so small that $|\epsilon_n| \, (< c_3\sqrt{\epsilon}) < \frac{1}{2}p^n$, then all the zeros of $f(w)$ lie in $|w| < (2c_4\sqrt{\epsilon})^{1/n}/p = c_5\epsilon^{1/2n}$. Of these zeros, n_1 belong to $f_1(w)$ and the remaining n_2 to $f_2(w)$. Hence it follows (omitting a few simple and obvious steps) that

$$g_1(z) = (q+pz)^{n_1} + \sum_{r=0}^{n_1} \alpha_r z^r$$

$$g_2(z) = (q+pz)^{n_2} + \sum_{r=0}^{n_2} \beta_r z^r$$

where every $|\alpha_r|$ and $|\beta_r|$ is less than $c_6\epsilon^{1/2n}$. Hence it follows that

$$|G_j(x) - B(x; n_j, p)| < c_7\epsilon^{1/2n} \qquad \dots \quad (3.30)$$

where $G_j(x)$ is the distribution function of Y_j $(j = 1, 2)$.

We now invoke a simple lemma due to Sapogov (1951).

Lemma : *If $H_1(x)$ and $H_2(x)$ are the distribution functions of the random variables Z_1 and Z_2 (not necessarily independent), then for all real x*

$$|H_1(x) - H_2(x)| \leqslant P(Z_1 \neq Z_2).$$

Proof : The event $(Z_1 > x) \supset (Z_2 > x, Z_1 - Z_2 \geqslant 0)$ so that the event $(Z_1 \leqslant x) \subset (Z_2 \leqslant x) \cup (Z_1 - Z_2 < 0)$,

and so

$$P(Z_1 \leqslant x) \leqslant P(Z_2 \leqslant x) + P(Z_1 - Z_2 < 0),$$

that is,

$$H_1(x) - H_2(x) \leqslant P(Z_1 - Z_2 < 0) \leqslant P(Z_1 \neq Z_2).$$

This relation, and the dual relation obtained from it by interchanging the subscripts, together yield the statement of the lemma.

In view of the above lemma and of the argument leading to relation (3.25), we immediately have, for $j = 1, 2$,

$$|F_j(x) - G_j(x)| \leqslant P(X_j \neq Y_j) < c_2\sqrt{\epsilon}. \qquad \dots \quad (3.31)$$

Relations (3.30) and (3.31) yield the assertion of our theorem for the case $c = 0$, where c is defined by (3.5). As we have already pointed out, it is sufficient to consider this case, since the general case can be reduced to this case by considering the random variables $X_1 - c$ and $X_2 + c$ instead of X_1 and X_2 respectively and applying the above argument to these new variables.

REFERENCES

LUKACS, E. (1960) : *Characteristic Functions*, Charles Griffin and Company, London.

SAPOGOV, N. A. (1951) : The problem of stability for Cramér's theorem (in Russian). *Izv. Akad. Nauk., SSSR, Ser. Matem.*, **13**, Number 3, 205-218.

——— (1959) : On the independent components of a sum of random variables which is distributed approximately Normally (in Russian). *Vestnik Leningrad Univ.*, **19**, 78-105.

SHALAEVSKY, O. V. (1959) : On stability for the theorem of D. A. Raikov (in Russian). *Vestnik Leningrad Univ.*, **7**, 41-49.

TITCHMARSH, E. C. (1939) : *The Theory of Functions*, Second Edition, Oxford University Press, Oxford.

Paper received : June, 1962.

THE LIMITING DISTRIBUTION OF THE VIRTUAL WAITING TIME AND THE QUEUE SIZE FOR A SINGLE-SERVER QUEUE WITH RECURRENT INPUT AND GENERAL SERVICE TIMES*

By LAJOS TAKÁCS

Columbia University, New York

SUMMARY. A single-server queueing process with recurrent input and general service times is considered and the limiting distribution of the virtual waiting time and that of the queue size are determined.

1. Introduction

Suppose that in the time interval $(0, \infty)$ customers arrive at a counter at times $\tau_1, \tau_2, \ldots, \tau_n, \ldots$ The customers are served by a single server in order of arrival. The server is idle if and only if there is no customer in the system. Denote by χ_n the service time of the n-th customer. It is supposed that the service times χ_n $(n = 1, 2, \ldots)$ and the interarrival times $\theta_n = \tau_{n+1} - \tau_n$ $(n = 1, 2, \ldots)$ are independent sequences of identically distributed, mutually independent, positive random variables with distribution functions

$$\boldsymbol{P}\{\chi_n \leqslant x\} = H(x) \qquad \ldots \quad (1.1)$$

and

$$\boldsymbol{P}\{\theta_n \leqslant x\} = F(x). \qquad \ldots \quad (1.2)$$

Let $\boldsymbol{E}\{\chi_n\} = \alpha$ and $\boldsymbol{E}\{\theta_n\} = \beta$. Throughout this paper α and β are supposed to be finite and the trivial case $\boldsymbol{P}\{\chi_n = \theta_n\} = 1$ is excluded.

Denote by $\eta(t)$ the virtual waiting time at time t, i.e., $\eta(t)$ is the time that a customer would have to wait if he arrived at time t. Let $\eta_n = \eta(\tau_n - 0)$, i.e., η_n is the actual waiting time of the n-th arriving customer. Denote by $\xi(t)$ the queue size at time t, i.e., the total number of customers (either waiting or being served) in the system at time t. Let $\xi_n = \xi(\tau_n - 0)$, i.e., ξ_n is the queue size immediately before the arrival of the n-th customer.

In what follows we shall determine the limiting distribution of $\eta(t)$ and that of $\xi(t)$ as $t \to \infty$. We note here that the distribution of the queue size is independent of the order of service.

At this point I should like to mention briefly the idea which leads to the notion of virtual waiting time. One can suppose, without loss of generality, that each customer is assigned his service time in advance at his arrival, because the service times are identically distributed, mutually independent random variables and independent of the arrival times. Suppose that we use a reading-timer which has a clock mechanism and each time a customer arrives we set the hand forward by his future service time. Since this clock runs as long as there are customers in the system, it will at any given instant show the appropriate virtual waiting time. Thus an arriving customer can

*This research was sponsored by the Office of Naval Research under Contract Number Nonr—266 (59).

immediately see his own actual waiting time on this clock. $\eta(t)$ can also be interpreted as the occupation time of the server at time t, that is, the time that is needed to complete the service of all those customers who arrived before t. In certain queues $\eta(t)$ has a real physical meaning. For instance, if we consider reading of messages in a telegraph office, then $\eta(t)$ can be interpreted as the length of all messages which remain to be read at time t.

The process $\{\eta(t)\}$ has interest not only in the theory of queues but also in the investigation of operation of dams. If $\eta(t)$ denotes the content of a dam at time t, then $\{\eta(t)\}$ has the same stochastic behaviour as the virtual waiting time in a queueing process (cf. Gani and Prabhu, 1959).

Finally, we introduce the following Laplace-Stieltjes transforms :

$$\psi(s) = \int_0^\infty e^{-sx}\, dH(x) \quad \ldots \quad (1.3)$$

and

$$\phi(s) = \int_0^\infty e^{-sx}\, dF(x) \quad \ldots \quad (1.4)$$

which are convergent if $\Re(s) \geqslant 0$.

2. The limiting distribution of the actual waiting time

The following results have been proved by Lindley (1952) : If $\alpha < \beta$, then the limiting distribution $\lim\limits_{n\to\infty} \boldsymbol{P}\{\eta_n \leqslant x\} = W(x)$ exists, independent of the initial state and it is the unique solution of the following integral equation of Wiener-Hopf type

$$W(x) = \begin{cases} \int_0^\infty K(x-y)\, dW(y) & \text{if } x \geqslant 0, \\ 0 & \text{if } x < 0 \end{cases} \quad \ldots \quad (2.1)$$

where

$$K(x) = \int_0^\infty H(x+y)\, dF(y) \quad \ldots \quad (2.2)$$

and further $W(0) > 0$. If $\alpha \geqslant \beta$ (the trivial case $\boldsymbol{P}\{\chi_n = \theta_n\} = 1$ is excluded), then $\boldsymbol{P}\{\lim\limits_{n\to\infty} \eta_n = \infty\} = 1$, whence it follows that $\lim\limits_{n\to\infty} \boldsymbol{P}\{\eta_n \leqslant x\} = 0$ for every x irrespective of the initial state.

Define the event $\mathcal{E}$ such that $\mathcal{E}$ is said to occur at the n-th arrival if the server is found to be idle at that time. Evidently, $\mathcal{E}$ is a recurrent event. If $\alpha \leqslant \beta$, then $\mathcal{E}$ is persistent, and if $\alpha > \beta$, then $\mathcal{E}$ is transient. (As to the theory of recurrent events we refer to Feller (1957), pp. 278-320.)

Denote by $R(x)$ the probability that the distance between two successive occurrences of $\mathcal{E}$ is $\leqslant x$. If $\alpha \leqslant \beta$, then $R(\infty) = 1$, i.e., $R(x)$ is a proper distribution function. The mean recurrence time of $\mathcal{E}$ is

$$\rho = \int_0^\infty x\, dR(x) = \int_0^\infty [1-R(x)]dx = \beta / W(0). \quad \ldots \quad (2.3)$$

If $\alpha < \beta$, then $\rho < \infty$, whereas if $\alpha = \beta$, then $\rho = \infty$. If $\alpha > \beta$, then $R(\infty) < 1$.

Finally, we note that if $F(x)$ is not a lattice distribution function, then $R(x)$ is not one either.

3. The limiting distribution of the virtual waiting time

We shall prove the following theorem.

Theorem 1 : *If $\alpha < \beta$ and $F(x)$ is not a lattice distribution function, then the limiting distribution*

$$\lim_{t\to\infty} \boldsymbol{P}\{\eta(t) \leqslant x\} = W^*(x) \qquad \dots \quad (3.1)$$

exists, independent of the initial state and is given by

$$W^*(x) = \left(1-\frac{\alpha}{\beta}\right)+\frac{\alpha}{\beta}W(x)*H^*(x) \qquad \dots \quad (3.2)$$

where $W(x)$ is defined by (2.1),

$$H^*(x) = \begin{cases} \dfrac{1}{\alpha}\displaystyle\int_0^x [1-H(y)]dy & \text{if } \ x \geqslant 0, \\ 0 & \text{if } \ x < 0, \end{cases} \qquad \dots \quad (3.3)$$

and $*$ denotes convolution. If $\alpha \geqslant \beta$ (the trivial case $\boldsymbol{P}\{\chi_n = \theta_n\} = 1$ is excluded), then $\lim_{t\to\infty} \boldsymbol{P}\{\eta(t) \leqslant x\} = 0$ for every x, irrespective of the initial state.

Proof : The proof consists of two parts. First we prove that the limit exists and then we find the explicit form of the limiting distribution in case of $\alpha < \beta$. We need the following lemma.

Lemma 1 : *Let A be an event which has the following property : If A occurs at time u and does not occur at time $u+t$, then this implies that at least one customer arrives in the interval $(u, u+t]$. Denote by $P_A(t)$ the probability that the system is in state A at time t. If $\alpha < \beta$ and $F(x)$ is not a lattice distribution function, then $\lim_{t\to\infty} P^*_A(t) = P^*_A$ exists and is independent of the initial state.*

Proof : Denote by $M(t)$ the expected number of occurrences of $\mathcal{E}$ in the time interval $(0, t]$. Let $Q^*_A(t)$ denote the probability that the system is in state A at time t and $\mathcal{E}$ never occurs in the interval $(0, t]$. Measuring time from an occurrence of $\mathcal{E}$ denote by $Q_A(t)$ the probability that A occurs at time t and $\mathcal{E}$ never occurs during the interval $(0, t]$. Evidently

$$P_A(t) = Q^*_A(t)+ \int_0^\infty Q_A(t-u)\, dM(u). \qquad \dots \quad (3.4)$$

If $\alpha < \beta$, then $\mathcal{E}$ is a persistent event and consequently $\lim_{t\to\infty} Q^*_A(t) = \lim_{t\to\infty} Q_A(t) = 0$. We whall prove later that $Q_A(u)$ is of bounded variation in every finite interval $[0, t]$. If $\alpha < \beta$ and $F(x)$ is not a lattice distribution function, then by a theorem of Blackwell (1948) we have for all $h > 0$ that

$$\lim_{t\to\infty} \frac{M(t+h)-M(t)}{h} = \frac{1}{\rho}, \qquad \dots \quad (3.5)$$

where ρ is defined by (2.3). Thus if $\alpha < \beta$ and $F(x)$ is not a lattice distribution function, then it follows from (3.4) and (3.5) that

$$\lim_{t\to\infty} P_A(t) = \frac{1}{\rho}\int_0^\infty Q_A(u)\,du \qquad \dots \quad (3.6)$$

irrespective of the initial state (cf. Smith (1958) and Takács (1962), pp. 227-228). Since

$$Q_A(u) \leqslant 1-R(u) \qquad \dots \quad (3.7)$$

for $u \geqslant 0$, the integral on the right hand side of (3.6) converges.

It remains to prove that $Q_A(u)$ is of bounded variation in any finite interval $[0, t]$. The proof is based on an idea of Smith (1955). Measure time from an occurrence of $\mathcal{E}$. Denote by $\nu(t)$ the number of arrivals in the interval $(0, t]$. Define $\chi_t = 1$ if A occurs at time t and $\mathcal{E}$ does not occur in the interval $(0, t]$; $\chi_t = 0$ otherwise. For $0 \leqslant u \leqslant t$ we have

$$|Q_A(t)-Q_A(u)| \leqslant \boldsymbol{E}\{\chi_t-\chi_u\}+2\boldsymbol{E}\{\nu(t)-\nu(u)\}. \qquad \dots \quad (3.8)$$

For, $\quad Q_A(t)-Q_A(u) = \boldsymbol{E}\{\chi_t-\chi_u\} = \boldsymbol{P}\{\chi_u = 0,\ \chi_t = 1\}-\boldsymbol{P}\{\chi_u = 1,\ \chi_t = 0\},$

whence $\quad |Q_A(t)-Q_A(u)| \leqslant \boldsymbol{E}\{\chi_t-\chi_u\}+2\boldsymbol{P}\{\chi_u = 1, \chi_t = 0\}$

and by assumption

$$\boldsymbol{P}\{\chi_u = 1, \chi_t = 0\} \leqslant \boldsymbol{P}\{\nu(t)-\nu(u) \geqslant 1\} \leqslant \boldsymbol{E}\{\nu(t)-\nu(u)\}.$$

Accordingly for any subdivision $0 = t_0 < t_1 < \dots < t_n = t$ of the interval $[0, t]$

$$\sum_{k=1}^{n} |Q_A(t_k)-Q_A(t_{k-1})| \leqslant \boldsymbol{E}\{\chi_t-\chi_0\}+2\boldsymbol{E}\{\nu(t)\} \leqslant 1+2\boldsymbol{E}\{\nu(t)\}. \qquad \dots \quad (3.9)$$

Since $\boldsymbol{E}\{\nu(t)\}$ is finite for every $t \geqslant 0$ ($\boldsymbol{E}\{\nu(t)\} \leqslant Ct+1$ where C is a positive constant), it follows that $Q_A(t)$ is of bounded variation in $[0, t]$. This completes the proof of Lemma 1.

Remark 1 : Let $\alpha < \beta$ and $F(x)$ be a non-lattice distribution function. Consider a monotone non-decreasing sequence of events $A_1 \subset A_2 \subset \dots \subset A_k \subset \dots$ for which A_k satisfies the assumptions of Lemma 1 and $\lim_{k\to\infty} A_k = \Omega$, the sure event. Let $\lim_{t\to\infty} P_{A_k}(t) = P^*_{A_k}$ defined by (3.6). Then $\lim_{k\to\infty} P^*_{A_k} = 1$. For, in this case $Q_{A_k}(u)$ $(k = 1, 2, \dots)$ is a monotone non-decreasing sequence and $\lim_{k\to\infty} Q_{A_k}(u) = 1-R(u)$, whence by Beppo Levi's theorem (cf. Riesz and Sz-Nagy (1952), p. 34)

$$\lim_{k\to\infty} \int_0^\infty Q_{A_k}(u)du = \int_0^\infty [1-R(u)]du = \rho.$$

The statement follows from (3.6).

Now we shall prove that if $\alpha < \beta$ and $F(x)$ is not a lattice distribution function, then the limit (3.1) exists and is independent of the initial state. Define A as the event that the virtual waiting time is $\leqslant x$, where $x \geqslant 0$. Then the event A satisfies the assumptions of Lemma 1 and $P_A(t) = \boldsymbol{P}\{\eta(t) \leqslant x\}$. Thus by Lemma 1

the limit (3.1) exists and $W^*(x)$ is a monotone non-decreasing function of x and by Remark 1 $W^*(\infty) = 1$.

If $\alpha \geqslant \beta$, then $\lim\limits_{t\to\infty} \boldsymbol{P}\{\eta(t) \leqslant x\} = 0$ for every x irrespective of the initial state. Since $\eta(t) \geqslant \eta_n$ for $\tau_{n-1} < t < \tau_n$, $n = 1, 2, \ldots$, and by Lindley's theorem $\boldsymbol{P}\{\lim\limits_{n\to\infty} \eta_n = \infty\} = 1$ for $\alpha \geqslant \beta$, we can conclude also that in this case $\boldsymbol{P}\{\lim\limits_{t\to\infty} \eta(t) = \infty\} = 1$.

Remark 2: Now we shall prove directly that $\boldsymbol{P}\{\lim\limits_{t\to\infty} \eta(t) = \infty\} = 1$ if $\alpha > \beta$. Denote by $\nu(t)$ the number of arrivals in the interval $(0, t]$. By a theorem of Doob (1948) we have

$$\boldsymbol{P}\left\{\lim_{t\to\infty} \frac{\nu(t)}{t} = \frac{1}{\beta}\right\} = 1. \qquad \ldots \quad (3.10)$$

Since obviously
$$\eta(t) \geqslant \eta(0) + \sum_{i=1}^{\nu(t)} \chi_i - t,$$

we have
$$\frac{\eta(t)}{t} \geqslant \frac{\eta(0)}{t} + \frac{\nu(t)}{t} \; \frac{1}{\nu(t)} \sum_{i=1}^{\nu(t)} \chi_i - 1. \qquad \ldots \quad (3.11)$$

If $t\to\infty$ in (3.11), then we have with probability one that $\eta(0)/t \to 0$, $\nu(t)/t \to 1/\beta$ and
$$\frac{1}{\nu(t)} \sum_{i=1}^{\nu(t)} \chi_i \to \alpha.$$

The latter follows from an easy extension of the strong law of large numbers. Thus by (3.11)
$$\liminf_{t\to\infty} \frac{\eta(t)}{t} \geqslant \frac{\alpha}{\beta} - 1 > 0$$

with probability one, whence
$$\boldsymbol{P}\{\lim_{t\to\infty} \eta(t) = \infty\} = 1. \qquad \ldots \quad (3.12)$$

This proves that if $\alpha > \beta$, then $\lim\limits_{t\to\infty} \boldsymbol{P}\{\eta(t) \leqslant x\} = 0$ for every x irrespective of the initial state.

Finally, it remains only to find $\lim\limits_{t\to\infty} \boldsymbol{P}\{\eta(t) \leqslant x\} = W^*(x)$ if $\alpha < \beta$ and $F(x)$ is not a lattice distribution function. First we define a random variable $\theta(t)$ as the time between t and the first arrival after t. Then we observe that the vector variables $\{\eta(t), \theta(t)\}$ form a Markov process. The initial state is given by $(\eta(0), \theta(0))$ where $\eta(0)$ is the initial occupation time of the server and $\theta(0) = \tau_1$ is the time of the first arrival. (We note that if the input is a Poisson process, then $\{\eta(t)\}$ is a Markov process in itself.) Define now A as follows: A occurs at time t if $\eta(t) \leqslant x$ and $\theta(t) \leqslant y$ where $x \geqslant 0$ and $y \geqslant 0$. This A satisfies the assumptions of Lemma 1 and if $\alpha < \beta$ and $F(x)$ is not a lattice distribution function, then by Lemma 1 we can conclude that

$$\lim_{t\to\infty} \boldsymbol{P}\{\eta(t) \leqslant x, \;\; \theta(t) \leqslant y\} = W^*(x, y) \qquad \ldots \quad (3.13)$$

exists and is independent of the initial state. $W^*(x, y)$ is a two dimensional distribution function, because by Remark 1 $W^*(\infty, \infty) = 1$. Let

$$\Omega^*(s, w, t) = \boldsymbol{E}\{e^{-s\eta(t)-w\theta(t)}\} \quad \ldots \quad (3.14)$$

and

$$\Omega^*(s, w) = \int_0^\infty\int_0^\infty e^{-sx-wy}d^2_{xy}W^*(x, y). \quad \ldots \quad (3.15)$$

If $\alpha < \beta$ and $F(x)$ is not a lattice distribution function, then by (3.13)

$$\lim_{t\to\infty} \Omega^*(s, w, t) = \Omega^*(s, w) \quad \ldots \quad (3.16)$$

for $\Re(s) \geqslant 0$ and $\Re(w) \geqslant 0$.

If

$$\Omega^*(s) = \int_0^\infty e^{-sx}dW^*(x), \quad \ldots \quad (3.17)$$

then obviously $\qquad \Omega^*(s) = \Omega^*(s, 0).$

Now we shall prove the following lemma.

Lemma 2 : *Denote by $m(t)$ the expected number of arrivals in the time interval* $(0, t]$. *If* $m(t+\Delta t)-m(t) = O(\Delta t)$,

then

$$\frac{\Omega^*(s, w, t+\Delta t)-\Omega^*(s, w, t)}{\Delta t} = (w+s)\Omega^*(s, w, t)-sP_0(t)\,\Phi\,(w, t)$$

$$-\frac{m(t+\Delta t)-m(t)}{\Delta t}[1-\psi(s)\phi(w)]\Omega(s, t)+\frac{o(\Delta t)}{\Delta t} \quad \ldots \quad (3.18)$$

where $\quad P_0(t) = \boldsymbol{P}\{\eta(t) = 0\}$, $\Omega\,(s, t) = \boldsymbol{E}\{e^{-s\eta(t)}\,|\,\theta(t) = 0\}$

and $\quad \Phi(w, t) = \boldsymbol{E}\{e^{-w\theta(t)}\,|\,\eta(t) = 0\}$.

Proof : If $\theta(t) > \Delta t$, then $\theta(t+\Delta t) = \theta(t)-\Delta t$ and $\eta(t+\Delta t) = \max(0, \eta(t)-\Delta t)$.

Thus

$$\boldsymbol{E}\{e^{-s\eta(t+\Delta t)-w\theta(t+\Delta t)}\,|\,\theta(t) > \Delta t\}$$
$$=\boldsymbol{P}\{\eta(t) = 0, \theta(t) > \Delta t\}\boldsymbol{E}\{e^{-s\eta(t)-w\theta(t)}\,|\,\eta(t) = 0, \theta(t) > \Delta t\}$$
$$(1+w\Delta t)+\boldsymbol{P}\{0 < \eta(t) \leqslant \Delta t, \theta(t) > \Delta t\}$$
$$\boldsymbol{E}\{e^{-s\eta(t)-w\theta(t)}\,|\,\eta(t) > \Delta t, \theta(t) > \Delta t\}[1+(w+s)\Delta t]+o(\Delta t).$$

If $\theta(t) \leqslant \Delta t$, then $\theta(t+\Delta t) = \theta-\epsilon_1\Delta t$ and $\eta(t+\Delta t) = \eta(t)+\chi-\epsilon_2\Delta t$ where χ is the total service time of all those customers who arrive in the interval $(t, t+\Delta t]$, θ is the interarrival time between the last arrival in $(t, t+\Delta t]$ and the first arrival after $t+\Delta t$, and further $0 \leqslant \epsilon_1 \leqslant 1$ and $0 \leqslant \epsilon_2 \leqslant 1$. Thus

$$\boldsymbol{E}\{e^{-s\eta(t+\Delta t)-w\theta(t+\Delta t)}\,|\,\theta(t) \leqslant \Delta t\} = \psi(s)\phi(w)\;\boldsymbol{E}\{e^{-s\eta(t)-w\theta(t)}\,|\,\theta(t) \leqslant \Delta t\}+0(\Delta t).$$

Since $\boldsymbol{P}\{\theta(t) \leqslant \Delta t\} \leqslant m(t+\Delta t)-m(t) = O(\Delta t)$ we obtain by the theorem of total expectation that

$$\Omega^*(s, w, t+\Delta t) = [1+(w+s)\Delta t]\Omega^*(s, w, t)-s\Delta t\,\boldsymbol{P}\{\eta(t) = 0, \theta(t) > \Delta t\}.$$
$$\boldsymbol{E}\{e^{-w\theta(t)}\,|\,\eta(t) = 0, \theta(t) > \Delta t\}-[1-\psi(s)\phi(w)]\;\boldsymbol{E}\{e^{-s\eta(t)}\,|\,\theta(t) \leqslant \Delta t\}.$$
$$\boldsymbol{P}\{\theta(t) \leqslant \Delta t\}+o(\Delta t). \quad \ldots \quad (3.19)$$

Since $m(t+\Delta t)-m(t) = \boldsymbol{P}\{\theta(t) \leqslant \Delta t\}+o(\Delta t)$ also holds, we get finally

$$\Omega^*(s, w, t+\Delta t) = [1+(w+s)\Delta t]\Omega^*(s, w, t)-s\Delta t\,\boldsymbol{P}\{\eta(t) = 0\}.$$
$$\boldsymbol{E}\{e^{-w\theta(t)}\,|\,\eta(t) = 0\}-[1-\psi(s)\phi(w)]\;\boldsymbol{E}\{e^{-s\eta(t)}\,|\,\theta(t) = 0\}.$$
$$[m(t+\Delta t)-m(t)]+o(\Delta t), \quad \ldots \quad (3.20)$$

which is in agreement with (3.18).

By Lemma 1 the following limits exist $\lim_{t\to\infty} \Omega^*(s, w, t) = \Omega^*(s, w)$ (cf.(3.16)], $\lim_{t\to\infty} P_0(t) = P_0^*$ (It is easy to prove directly that $P_0^* = 1-\alpha/\beta$ [cf. Takács (1962) p. 142]) and $\lim_{t\to\infty} \Phi(w, t) = \Phi(w)$, say. By Lindley's theorem $\lim_{t\to\infty} \Omega(s, t) = \Omega(s)$

where
$$\Omega(s) = \int_0^\infty e^{-sx}\, dW(x) \qquad \dots \quad (3.21)$$

and $W(x)$ is defined by (2.1). By Blackwell's theorem
$$\lim_{t\to\infty} \frac{m(t+\Delta t)-m(t)}{\Delta t} = \frac{1}{\beta}. \qquad \dots \quad (3.22)$$

If we let $t\to\infty$ in (3.18) we obtain
$$(w+s)\Omega^*(s, w) = sP_0^*\Phi(w) + \frac{[1-\psi(s)\phi(w)]}{\beta}\,\Omega(s). \qquad \dots \quad (3.23)$$

If $w\to 0$ in (3.23), then we get
$$\Omega^*(s) = P_0^* + \frac{[1-\psi(s)]}{\beta s}\,\Omega(s). \qquad \dots \quad (3.24)$$

Since $\Omega^*(0) = 1$, we obtain that $P_0^* = 1-\alpha/\beta$. Thus finally the Laplace-Stieltjes transform of $W^*(x)$ is given by
$$\Omega^*(s) = \left(1-\frac{\alpha}{\beta}\right) + \frac{\alpha}{\beta}\,\frac{[1-\psi(s)]}{\alpha s}\,\Omega(s), \qquad \dots \quad (3.25)$$

whence (3.2) follows by inversion. This completes the proof of Theorem 1.

Examples : (i) Suppose that $F(x) = 1-e^{-\lambda x}(x \geqslant 0)$ and $H(x)$ is arbitrary. In this case $\beta = 1/\lambda$. If $\lambda\alpha < 1$, then
$$\Omega(s) = \frac{1-\lambda\alpha}{1-\lambda\dfrac{1-\psi(s)}{s}} \qquad \dots \quad (3.26)$$

and thus by (3.25) $\Omega^*(s) = \Omega(s)$, i.e., $W^*(x) = W(x)$.

(ii) Suppose that $F(x)$ is arbitrary and $H(x) = 1-e^{-\mu x}\,(x \geqslant 0)$. In this case $\alpha = 1/\mu$. If $\mu\beta > 1$, then
$$\Omega(s) = (1-\delta)+\delta\,\frac{\mu(1-\delta)}{\mu(1-\delta)+s} \qquad \dots \quad (3.27)$$

where $z = \delta$ is the only root of $z = \phi(\mu(1-z))$ inside the unit circle. Now if we suppose that $F(x)$ is not a lattice distribution function, then we obtain by (3.25) that
$$\Omega^*(s) = \left(1-\frac{1}{\mu\beta}\right) + \frac{1}{\mu\beta}\,\frac{\mu(1-\delta)}{\mu(1-\delta)+s}. \qquad \dots \quad (3.28)$$

From (3.27)
$$W(x) = 1-\delta e^{-\mu(1-\delta)x} \quad \text{if } x \geqslant 0, \qquad \dots \quad (3.29)$$

and from (3.28)
$$W^*(x) = 1-\frac{1}{\mu\beta}\,e^{-\mu(1-\delta)x} \quad \text{if } x \geqslant 0. \qquad \dots \quad (3.30)$$

4. THE LIMITING DISTRIBUTION OF THE QUEUE SIZE

The following two theorems are concerned with the limiting distribution of the queue size. Formulas (4.2) and (4.10) were first found by Mr. M. Aczél (oral communication made in January 1958 during a meeting on "Queueing Theory and Practice" arranged by the Institute for Engineering Production of the University of Birmingham, England). Formula (4.10) has also been proved by Kawata (1961). Under some restrictive conditions the existence of the limiting distribution of the queue size for many-server queues has been investigated by Finch (1959).

Theorem 2 : *If $\alpha < \beta$ and $F(x)$ is not a lattice distribution function, then the limiting distribution $\lim_{t\to\infty} \boldsymbol{P}\{\xi(t) = k\} = P_k^*$ $(k = 0, 1, \ldots)$ exists and is independent of the intitial queue size. We have*

$$P_0^* = 1 - \frac{\alpha}{\beta} \qquad \ldots \quad (4.1)$$

and for $k = 1, 2, \ldots$
$$P_k^* = \frac{\alpha}{\beta}\int_0^\infty [F_{k-1}(x) - F_k(x)]\, d[W(x) * H^*(x)] \qquad \ldots \quad (4.2)$$

where $F_k(x)$ denotes the k-th iterated convolution of $F(x)$ with itself; $F_0(x) = 1$ if $x \geqslant 0$, and $F_0(x) = 0$ if $x < 0$; $W(x)$ is defined by (2.1) $H^*(x)$ is defined by (3.3). If $\alpha \geqslant \beta$, then $\lim_{t\to\infty} \boldsymbol{P}\{\xi(t) = k\} = 0$ $(k = 0, 1, \ldots)$ irrespective of the initial queue size.

Proof: If we define A as the event that the queue size is $\leqslant k$ $(k = 0, 1, \ldots)$, then this event satisfies the conditions of Lemma 1. Accordingly if $\alpha < \beta$ and $F(x)$ is not a lattice distribution function, then the limiting distribution $\lim_{t\to\infty} \boldsymbol{P}\{\xi(t) \leqslant k\}$ exists and is independent of the initial state. If $\alpha \geqslant \beta$, then $\lim_{t\to\infty} \boldsymbol{P}\{\xi(t) \leqslant k\} = 0$ $(k = 0, 1, \ldots)$. This follows from (4.13) which will be proved later.

Remark 3 : We shall give a direct proof for the case $\alpha > \beta$. Denote by $\delta(t)$ the number of departures in the time interval $(0, t]$ and by $\nu(t)$ the number of arrivals in $(0, t]$. Then $\xi(t) = \xi(0) + \nu(t) - \delta(t)$,

whence
$$\frac{\xi(t)}{t} = \frac{\xi(0)}{t} + \frac{\nu(t)}{t} - \frac{\delta(t)}{t}. \qquad \ldots \quad (4.3)$$

By Doob's theorem $\lim_{t\to\infty} \nu(t)/t = 1/\beta$ and $\limsup_{t\to\infty} \delta(t)/t \leqslant 1/\alpha$ with probability one. Since $\lim_{t\to\infty} \xi(0)/t = 0$ with probability 1 we obtain from (4.3) that

$$\liminf_{t\to\infty} \frac{\xi(t)}{t} \geqslant \frac{1}{\beta} - \frac{1}{\alpha} > 0$$

with probability one, i.e.,
$$\boldsymbol{P}\{\lim_{t\to\infty} \xi(t) = \infty\} = 1. \qquad \ldots \quad (4.4)$$

To find P_k^* for $k = 1, 2, \ldots$ we can write that

$$P\{\xi(t) = k\} = Q_k^*(t) + \int_0^t [F_{k-1}(t-u) - F_k(t-u)] \int_0^{t-u} [1 - H(t-u-y)].$$

$$d_y P\{\eta(u) \leqslant y \mid \theta(u) = 0\} d\, m(u), \qquad \ldots \quad (4.5)$$

where $Q_k^*(t)$ is the probability that the queue size is k at time t and there is no arrival in the time interval $(0, t]$. The second term on the right hand side of (4.5) can be obtained in the following way: The customer being served at time t arrives at time u $(0 \leqslant u \leqslant t)$, his waiting time is y $(0 \leqslant y \leqslant t-u)$ and in the interval $(u, t]$ $k-1$ customers arrive. If $\alpha < \beta$ and $F(x)$ is not a lattice distribution function, then $\lim_{u \to \infty} P\{\eta(u) \leqslant y \mid \theta(u) = 0\} = W(y)$ where $W(y)$ is defined by (2.1). By using (3.22) we obtain from (4.5) that $\lim_{t \to \infty} P\{\xi(t) = k\} = P_k^*$ where P_k^* is given by (4.2) for $k = 1, 2, \ldots$. If $k = 0$, then $P_k^* = 1 - \alpha/\beta$, because

$$\lim_{t \to \infty} P\ \{\xi(t) = 0\} = 1 - \sum_{k=1}^{\infty} P_k^* = 1 - \frac{\alpha}{\beta}. \qquad \ldots \quad (4.6)$$

Remark 4: If

$$W_1 = \int_0^{\infty} x\, dW(x) \qquad \ldots \quad (4.7)$$

is finite, then

$$\sum_{k=0}^{\infty} kP_k^* = \frac{1}{\beta}\, (W_1 + \alpha). \qquad \ldots \quad (4.8)$$

An intuitive proof is as follows: Obviously

$$\int_0^t \xi(u)\, du - \sum_{i=1}^{\nu(t)} (\eta_i + \chi_i)$$

is bounded with probability 1, i.e.,

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t \xi(u) du = \lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{\nu(t)} (\eta_i + \chi_i) = \frac{1}{\beta} (W_1 + \alpha) \qquad \ldots \quad (4.9)$$

with probability one. For, by (3.10) $\nu(t)/t \to 1/\beta$ with probability 1, by the strong law of large numbers

$$\lim_{t \to \infty} \frac{1}{\nu(t)} \sum_{i=1}^{\nu(t)} \chi_i = \alpha$$

with probability 1, and by the ergodic theorem

$$\lim_{t \to \infty} \frac{1}{\nu(t)} \sum_{i=1}^{\nu(t)} \eta_i = W_1$$

with probability 1. If we suppose that $\{\xi(t)\}$ is a stationary process, then evidently

$$E\{\xi(t)\} = \sum_{k=0}^{\infty} kP_k^*$$

for every $t \geqslant 0$, and if we form the expectation of (4.9) we obtain (4.8).

Theorem 3 : *If* $\alpha < \beta$, *then* $\lim\limits_{n\to\infty} \boldsymbol{P}\{\xi_n = k\} = P_k$ $(k = 0, 1, \ldots)$ *exists, independent of the initial queue size and we have*

$$P_k = \int_0^\infty [F_k(x) - F_{k+1}(x)]d[W(x)*H(x)] \qquad \ldots \quad (4.10)$$

where $W(x)$ *is defined by* (2.1). *If* $\alpha \geqslant \beta$, *then* $\lim\limits_{n\to\infty} P\{\xi_n = k\} = 0$ *for every* k *irrespective of the initial queue size.*

Proof : The event $\xi_{n+k+1} \leqslant k$ occurs if and only if the n-th arriving customer departs before the $n+k+1$-st customer arrives, i.e., if and only if the queue size immediately after the departure of the n-th arriving customer is $\leqslant k$. Thus for arbitrary initial queue size $\xi(0)$ we have

$$\boldsymbol{P}\{\xi_{n+k+1} \leqslant k\} = \int_0^\infty [1 - F_{k+1}(x)]d[W_n(x)*H(x)] \qquad \ldots \quad (4.11)$$

where $W_n(x) = \boldsymbol{P}\{\eta_n \leqslant x\}$, because the queue size immediately after the departure of the n-th arriving customer is equal to the number of arrivals during the waiting time and the service time of the n-th arriving customer.

If $\alpha < \beta$, then $\lim\limits_{n\to\infty} W_n(x) = W(x)$ and by (4.11)

$$\lim_{n\to\infty} \boldsymbol{P}\{\xi_n \leqslant k\} = \int_0^\infty [1 - F_{k+1}(x)]d[W(x)*H(x)] \qquad \ldots \quad (4.12)$$

which proves (4.10).

If $\alpha \geqslant \beta$, then $\lim\limits_{n\to\infty} W_n(x) = 0$ for every x and by (4.11)

$$\lim_{n\to\infty} \boldsymbol{P}\{\xi_n \leqslant k\} = 0 \qquad \ldots \quad (4.13)$$

for every k.

References

BLACKWELL, D. (1948) : A renewal theorem. *Duke Math. Jour.*, **15**, 145-150.

DOOB, J. L. (1948) : Renewal theory from the point of view of the theory of probability. *Trans. Amer. Math. Soc.*, **63**, 422–438.

FELLER, W. (1957) : *An Introduction to Probability Theory and its Applications*, Vol. 1., Second edition, John Wiley, New York,

FINCH, P. D. (1959) : On the distribution of queue size in queueing problems. *Acta Math. Acad. Sci. Hungar*, **10**, 327–336.

GANI, J. and PRABHU, N. U. (1959) : The time-dependent solution for a storage model with Poisson input. *Journal of Mathematics and Mechanics*, **8**, 653–664.

LINDLEY, D. V. (1952) : The theory of queues with a single server. *Proc. Camb. Phil. Soc.*, **48**, 277–289.

KAWATA, T. (1961) : On the imbedded queueing process of general type. *Bull. Int. Stat. Inst.*, **38**, 445–455.

RIESZ, F. and SZ.-NAGY, B. (1952) : Lecons d'analyse fonctionnelle. *Akadémiai Kiadó*, Budapest.

SMITH, W. L. (1955) : Regenerative stochastic processes. *Proc. Roy. Soc.*, Series A, **232**, 6–31.

——— (1958) : Renewal theory and its ramifications. *J. Roy. Stat. Soc.*, Series B, **20**, 243–302.

TAKÁCS, L. (1962) : *Introduction to the Theory of Queues*, Oxford University Press, New York.

Paper received : December, 1962.

THE CAPACITY OF AN INDECOMPOSABLE CHANNEL*

By J. WOLFOWITZ

Cornell University

SUMMARY. The strong converse of the coding theorem for an indecomposable channel is proved, thus establishing the capacity of the channel. An approximating result is proved which makes the capacity computable.

1. Description of an indecomposable channel

Let $\bar{A} = \{1, \ldots, a\}$ and $\bar{B} = \{1, \ldots, b\}$ be, respectively, the input and output alphabets of an indecomposable channel. (What these terms are will be apparent shortly.) To avoid the trivial both a and b should be greater than one. A sequence of n "letters," each an element (integer) in A (respectively, in B) is called a transmitted (respectively, a received) n-sequence, also sometimes a word or an n-word. All words sent (i.e., transmitted) and received will have length n (i.e., consist of n letters), where n is an arbitrary integer.

An indecomposable channel is characterized by the way it transmits any transmitted n-sequence, and this can be described with the aid of a stochastic $b \times b$ matrices $D_1, \ldots, D_a$. The element of D_k in the i-th row and j-th column will be called $d(k, i, j)$, $k = 1, \ldots, a$; $i, j = 1, \ldots, b$, and the D's are subject to a restriction which will be described below. Let

$$u_0 = (x_1, \ldots, x_n)$$

be any transmitted n-sequence and

$$v_0 = (y_1, \ldots, y_n)$$

be any received n-sequence. When u_0 is sent or transmitted over the channel the n-sequence then received is a chance sequence (one whose components are chance variables), say

$$v(u_0) = (Y_1(u_0), \ldots, Y_n(u_0)).$$

The description of the channel will be complete when we have defined the probability that $v(u_0) = v_0$ (for any u_0 and v_0). This probability depends upon the "initial state" of the channel (i.e., the state when the transmission of u_0 begins), and this state is an element of $\bar{B}$, say h. Then we define, for $h = 1, \ldots, b$,

$$P\{v(u_0) = v_0 | h\} = d(x_1, h, y_1) \prod_{m=2}^{n} d(x_m, y_{m-1}, y_m). \quad \ldots \quad (1.1)$$

The specification (1.1) can be thought of as follows: At the beginning of the transmission of u_0 the channel is in state h, either because this was the last letter received in the previous transmission or for some other reason. The first letter sent is x_1. Employing the stochastic matrix D_{x_1} we have that the probability of the channel's

*Research supported by the Office of Scientific Research and Development of the U.S. Air Force.

moving from state h to state y_1 (i.e., of the letter's y_1 being received) is $d(x_1, h, y_1)$. After the movement the channel is in state y_1, and the procedure is repeated, this time starting from state y_1 and using D_{x_2}, because x_2 is the second letter sent. Successive movements take place in the same manner. The right member of (1.1) shows that, for given $Y_1(u_0), \ldots, Y_m(u_0)$, the conditional distribution of $Y_{m+1}(u_0)$ depends only on $Y_m(u_0)$ and the $(m+1)$-st element x_{m+1} of u_0.

A code (n, N, λ) for an indecomposable channel is a system

$$\{(u_1, A_1), \ldots, (u_N, A_N)\} \qquad \ldots \quad (1.2)$$

where $u_1, \ldots, u_N$ are transmitted n-sequences, $A_1, \ldots, A_N$ are *disjoint* sets of received n-sequences, and

$$P\{v(u_i) \epsilon A_i \,|\, h\} \geqslant 1-\lambda, \qquad \ldots \quad (1.3)$$

$$h = 1, \ldots, b \,; \; i = 1, \ldots, N.$$

The reason for the name "code" and the use to which a code can be put are obvious. (For more about codes see, for example, Wolfowitz (1961)). N is called the length of the code, λ is called the probability of error, and n is the length of each word.

A stochastic matrix is called indecomposable if, in the terminology of Doob (1953, p. 179), it contains only one ergodic class, or, in the terminology of Feller (1957, p. 355), it contains at most one closed set of states other than the set of all states. It is called aperiodic (ibid.) if it has period 1. Let D be a stochastic indecomposable aperiodic (SIA) matrix. Then it is proved in books on Markov chains [e.g., (Doob, 1953) and (Feller, 1957)] that D^n approaches, as $n \to \infty$, a stochastic matrix all of whose rows are identical, and conversely, if D is a stochastic matrix such that D^n approaches a (stochastic) matrix all of whose rows are identical then D is SIA. This property could be used to furnish an alternate definition of an SIA matrix.

A channel is called indecomposable if (and only if) every product of any number of the matrices D (with replications permitted, of course) is itself SIA. (Of course, then, each D, being itself such a product, must be SIA.) Indecomposable channels were introduced by Blackwell, Breiman, and Thomasian (1958). Thomasian gave a finite algorithm for deciding whether a given finite set of SIA matrices is such that every product of these matrices is itself SIA (see also Wolfowitz (1963)).

2. Coding theorems

Hereafter we assume that we are dealing with an indecomposable channel. All logarithms will be assumed to be to the base 2; this is due to a convention in information theory which has lost all significance but is innocuous. Whenever in what follows an algebraic expression formally appears to be 0 log 0 the latter value is to be considered 0.

Let l be any integer, and let h designate the initial state of the channel. Let

$$U = (X_1, \ldots, X_l)$$

be a chance transmitted l-sequence, i.e., $X_1, \ldots, X_l$ are chance variables (not in general independent) with values in $\bar{A}$. Let Q' be the distribution of U; i.e., if u' is any transmitted l-sequence then

$$Q'(u') = P\{U = u'\}.$$

Let V be the chance l-sequence received when U is transmitted (the initial state of the channel is h). Let Q be the joint distribution of (U, V) and Q'' be the (marginal) distribution of V. Thus, if v' is any received l-sequence, then

$$Q(u', v') = P\{U = u', \ V = v'\}$$
$$Q''(v') = P\{V = v'\} = \sum_{u'} Q(u', v').$$

Of course Q' determines both Q and Q'' (for a given channel). Define

$$I(Q', h) = \frac{1}{l} \sum_{u', v'} Q(u', v') \log \frac{Q(u', v')}{Q'(u')Q''(v')}. \quad \ldots \quad (2.1)$$

(Clearly, Q and Q'' depend upon h.) Finally, define

$$G(l) = \max_{Q'} \ \min_{h} \ I(Q', h) \quad \ldots \quad (2.2)$$

and

$$C = \sup_{l} \ G(l). \quad \ldots \quad (2.3)$$

The following theorems were proved by Blackwell, Breiman, and Thomasian (1958); see also Wolfowitz (1961, p. 75).

Theorem 1 : *Let $\epsilon > 0$ and λ, $0 < \lambda \leqslant 1$, be arbitrary. For n sufficiently large there exists a code*

$$(n, \ 2^{n(C-\varepsilon)}, \ \lambda)$$

for the indecomposable channel.

Theorem 2 : *For any n a code (n, N, λ) for the indecomposable channel satisfies*

$$\log N < \frac{nC+1}{1-\lambda}.$$

In the present paper we will prove

Theorem 3 : *Let $\epsilon > 0$ and λ, $0 \leqslant \lambda < 1$, be arbitrary. For all n sufficiently large there does not exist a code*

$$(n, 2^{n(C+\varepsilon)}, \lambda)$$

for the indecomposable channel.

Theorem 2 is a so-called weak converse. Theorem 3 is a so-called strong converse of the coding theorem. Theorems 1 and 3 justify us in calling C the "capacity" of the indecomposable channel (Wolfowitz, 1961, p. 59).

It seems desirable that the determination of the capacity of the channel should be such that it is possible, at least in principle, to compute the capacity to within any specified accuracy. The expression (2.3) does not meet this requirement, but Theorem 4, to be stated and proved in Section 3, does enable us to do this.

Proof of Theorem 3 : For any stochastic matrix $M = \{m_{ij}\}$, define

$$\delta(M) = \max_{j} \max_{i_1, i_2} |m_{i_1 j} - m_{i_2 j}|$$

Thus $\delta(M) = 0$ implies that all rows of M are identical, and $\delta(M)$ small implies that they are almost identical. Essential to our proof will be the following result, proved in Wolfowitz (1963).

Theorem A : *Let $M_1, \ldots, M_a$ be (finite, square) stochastic matrices such that the product of any number of M's (repetition permitted) is SIA. Let $\eta > 0$ be arbitrary. There exists an integer $r(\eta)$ such that every product R of at least $r(\eta)$ M's satisfies*

$$\delta(R) < \eta. \qquad \ldots \quad (2.4)$$

We now define

$$c(\eta) = \max_{\substack{|x-x'| = \eta \\ 0 \leqslant x < x' \leqslant 1}} |x \log x - x' \log x'|$$

and, for $\eta > 0$ and t a positive integer, let

$$z(\eta, t) = b\eta(\log a + 2 \log b) + \frac{2bc(\eta)}{t}$$

Let the (arbitrary) ϵ of Theorem 3 be given. Let r and t be positive integers which will be further described later. We will prove Theorem 3 for all sufficiently large n of the form $k(r+t)$, with k an integer. The proof for all sufficiently large n will then be trivial.

Let the system (1.2) be a code (n, N, λ) for the indecomposable channel, with $\lambda < 1$. Let v_0, say, be any sequence in any one of the A's, say A_i. To A_i add *all* received n-sequences which coincide with v_0 in the places whose serial numbers are of the form

$$k'(r+t)+j,$$

where k' is an integer and j takes the values $r+1, r+2, \ldots, r+t$. Perform this operation for all v_0 in some A_i, and designate A_i after addition of all sequences in this manner by A_i^*. It is clear that each sequence in A_i has in this manner been replaced by at most b^{kr} sequences. The system

$$\{(u_1, A_1^*), \ldots, (u_N, A_N^*)\}$$

would be a code (n, N, λ) if it were not for the fact that the A_i^* are not disjoint.

Let u_i be any transmitted n-sequence of the code (1.2), and let $v(u_i)$ be the chance received sequence. Let $k' \leqslant k-1$ be any integer, and let $t' = k'(r+t)+r$. When the $(t'+1)$-st element of u_i is being transmitted the channel is in state $Y_{t'}(u_i)$, whose distribution depends upon the first t' elements of u_i and the state of the channel at the beginning of the transmission of u_i. Equivalently, the distribution of $Y_{t'}(u_i)$ is determined by $Y_{k'(r+t)}(u_i)$ and the r elements of u_i which precede the $(t'+1)$-st.

Consider now the following channel K: Its input alphabet consists of sequences of $(r+t)$ letters from $\bar{A}$. Its output alphabet consists of sequences of t letters from $\bar{B}$. Let c (respectively d) be any element of the input (respectively output) alphabet of channel K. Define $P\{v(c) = d\}$ according to the channel K to be equal to the probability that, when the $(r+t)$-sequence c is sent over the original indecomposable channel, the last t elements of the received $(r+t)$-sequence are d. This probability depends upon the initial state of the channel at the beginning of transmission of each letter (of the input alphabet of channel K), i.e., except in the case of the *first* letter of a word, upon the previous received letter. Thus the channel has memory. However, we will regard it as a memoryless channel and leave an ambiguity in the above definition of the channel probability function. It will be shown that, under certain conditions which will prevail, this ambiguity will be of so little consequence that the desired result will be achieved. As an example of how a memoryless channel might be achieved we give the following: After the transmission of any (input) letter of the channel K the channel is considered to be restored to a given fixed state.

Suppose *temporarily* that the operation of the original indecomposable channel is changed *only for the first letter* (transmitted and received) as follows: There is given a stochastic $b \times b$ matrix $Z = \{z_{ij}\}$, such that $\delta(Z) < \eta$. If h is the initial state of the channel let the probability that the first received letter will be $j (j = 1, \ldots, b)$ be z_{hj}. After the first letter is received the channel behaves like the indecomposable channel which was originally defined. For this modified channel we will shortly prove that

$$\max_h I(Q', h) - \min_h I(Q', h) < lz(\eta, l) \qquad \ldots \quad (2.5)$$

independent of Z.

Assume temporarily that (2.5) is true. Let $r(\eta)$ be sufficiently large so that (2.4) holds with the M's replaced by $D_1, \ldots, D_a$. Then it follows, from (2.5), the choice of r, and the definition of the channel K, that the capacity of the latter is less than

$$\frac{t}{r+t}\,(G(t)+z(\eta, t)). \qquad \ldots \quad (2.6)$$

By essentially the same argument as in [Wolfowitz (1961), Section 5.4], we obtain from (2.6) that, for all k sufficiently large $(n = k(r+t))$,

$$N < b^{kr} \cdot 2^{kt(G(t)+\varepsilon/4+z(\eta, t))} \qquad \ldots \quad (2.7)$$

Now choose η so small that $z(\eta, t) < \epsilon/4$, and then t so large that

$$\frac{r \log b}{r+t} < \frac{\epsilon}{4}.$$

Theorem 3 then follows from (2.7) when n is of the form $k(r+t)$. The theorem is then obvious for general n, as in [Wolfowitz (1961), equation (5.4.6)].

It remains to prove (2.5). We consider

$$\sum_{u', v'} Q_h(u', v') \log Q_h(u', v') = F(h), \text{ say}, \quad \ldots \quad (2.8)$$

for two different initial states, h_1 and h_2. Now $Q_h(u', v')$ equals $Q'(u')$ (which does not depend on h), multiplied by the conditional probability $\alpha_h(u', m)$ of the first letter m (say) of v', multiplied by $\beta(u', v')$, the conditional (upon the two preceding events) probability of the sequence q (say) of the last $(l-1)$ letters of v' (which also does not depend on h). (Sometimes, when it will make the arguments clearer, we will write $\alpha_h(u', v')$ and $\beta(u', m, q)$.) From the definition of the modified indecomposable channel we have that

$$|\alpha_{h_1}(u', v') - \alpha_{h_2}(u', v')| < \eta. \quad \ldots \quad (2.9)$$

Hence

$$\begin{aligned}
&|F(h_1)-F(h_2)| \\
&= |\sum_{u', v'} Q_{h_1}(u', v') [\log (Q'(u')\, \beta(u', v')+\log \alpha_{h_1}(u', v')] \\
&\quad - \sum_{u', v'} Q_{h_2}(u', v') [\log(Q'(u')\, \beta(u', v'))+\log \alpha_{h_2}(u', v')]| \\
&< -\eta \sum_{u', v'} [Q'(u')\, \beta(u', v')] \log[Q'(u')\, \beta(u', v')] \\
&\quad + \sum_{u', v'} [Q'(u')\, \beta(u', v')]\, |\alpha_{h_1}(u', v') \log \alpha_{h_1}(u', v') \\
&\qquad\qquad - \alpha_{h_2}(u', v') \log \alpha_{h_2}(u', v')|. \quad \ldots \quad (2.10)
\end{aligned}$$

Also,

$$\begin{aligned}
&\sum_{u', v'} Q'(u')\, \beta(u', v') \quad \ldots \quad (2.11) \\
&= \sum_{m} \sum_{u'} \sum_{q} Q'(u')\, \beta(u', m, q) = \sum_{m} \sum_{u'} Q'(u') = b.
\end{aligned}$$

Hence the second term of the right member of (2.10) is not greater than $bc(\eta)$. Also,

$$\begin{aligned}
&-\sum_{u', v'} [Q'(u')\, \beta(u', v')] \log [Q'(u')\, \beta(u', v')] \quad \ldots \quad (2.12) \\
&= -\sum_{m} \sum_{u', q} [Q'(u')\, \beta(u', m, q)] \log [Q'(u')\, \beta(u', m, q)].
\end{aligned}$$

From (2.11) and, e.g., [Wolfowitz (1961), (2.2.4)], we obtain that the right member of (2.12) is not greater than

$$b \log (a^l b^{l-1}) < lb \log ab. \quad \ldots \quad (2.13)$$

Thus we have that the right member of (2.10) is less than

$$l\eta b \log ab+bc(\eta). \qquad \dots \quad (2.14)$$

The expression

$$\sum_{u',\, v'} Q_h(u', v') \log Q'(u') \qquad \dots \quad (2.15)$$

which enters into $I(Q', h)$ does not depend upon h. It remains therefore to consider

$$\begin{aligned} S(h) &= \sum_{u'.\, v'} Q_h(u',\, v') \log Q''_h(v') = \sum_{v'} (\sum_{u'} Q_h(u', v') \log \sum_{u'} Q_h(u', v')) \\ &= \sum_{m,q} (\sum_{u'} Q'(u')\alpha_h(u', m)\, \beta(u', m, q)) \log \sum_{u'} Q'(u')\, \alpha_h(u', m)\, \beta(u', m, q)). \end{aligned} \qquad \dots \quad (2.16)$$

Define $T_h(v')$ by

$$T_h(v') \sum_{u'} Q'(u')\, \beta(u', v') = \sum_{u'} Q_h(u', v'). \qquad \dots \quad (2.17)$$

Obviously, for all v' and h,

$$0 \leqslant T_h(v') \leqslant 1 \qquad \dots \quad (2.18)$$

and, from (2.9) and (2.17),

$$|T_{h_1}(v')-T_{h_2}(v')| < \eta. \qquad \dots \quad (2.19)$$

Thus

$$\begin{aligned} S(h) &= \sum_m \sum_q \{[T_h(v') \log T_h(v')] \sum_{u'} Q'(u')\, \beta(u', m, q)\} \\ &\quad + \sum_m \sum_q (T_h(v') \sum_{u'} Q'(u')\ \beta(u', m, q) \log \sum_{u'} Q'(u')\, \beta(u', m, q) \\ &= V_1(h)+V_2(h) \text{ (say).} \end{aligned} \qquad \dots \quad (2.20)$$

From (2.11), (2.19) and [Wolfowitz (1961)], (2.2.4), we obtain that

$$|V_2(h_1)-V_2(h_2)| < b\eta\, l \log b. \qquad \dots \quad (2.21)$$

From (2.11) and (2.18) we obtain that

$$|V_1(h_1)-V_1(h_2)| \leqslant bc(\eta). \qquad \dots \quad (2.22)$$

Finally, from (2.14), (2.21) and (2.22) we obtain (2.5). This completes the proof of Theorem 3.

3. Rapidity of approach to C : Theorem 4

Let $r^*(\eta)$ be the smallest integer such that (2.4) holds with the M's replaced by $D_1, \dots, D_a$. Let t be any positive integer. From (2.7) and Theorem 1 we obtain that, for n sufficiently large

$$C-\frac{5\epsilon}{4} < \frac{r^*(\eta) \log b+tz(\eta, t)+tG(t)}{t+r^*(\eta)}. \qquad \dots \quad (3.1)$$

But (3.1) does not involve n and ϵ was arbitrary. Hence, from (3.1) and the definition of $G(t)$ we have

$$G(t) \leqslant C \leqslant \frac{r^*(\eta) \log b+tz(\eta,t)+tG(t)}{t+r^*(\eta)}. \qquad \dots \quad (3.2)$$

When η is sufficiently small and t is (then) sufficiently large the expression

$$J(\eta, t) = \frac{r^*(\eta) \log b + tz(\eta, t)}{t + r^*(\eta)} \quad \ldots \quad (3.3)$$

can be made as small as desired. Thus we have proved

Theorem 4 : *Let C be the capacity of an indecomposable channel. Let $\eta > 0$ and the positive integer t both be arbitrary. Then*

$$C - G(t) \leqslant J(\eta, t) - \frac{r^*(\eta) G(t)}{t + r^*(\eta)}. \quad \ldots \quad (3.4)$$

For η sufficiently small and then t (depending on η) sufficiently large the right member of (3.4) can be made arbitrarily small.

Theorem 4 enables us, at least in principle, to compute C to any specified approximation.

REFERENCES

BLACKWELL, D., BREIMAN, L., and THOMASIAN, A. J. (1958) : Proof of Shannon's transmission theorem for finite-state indecomposable channels. *Ann. Math. Stat.*, **29**, 4, 1209-1220.

DOOB, J. L. (1953) : *Stochastic Processes*, John Wiley and Sons, New York.

FELLER, W. (1957) : *An Introduction to Probability Theory and its Applications*, Vol. 1, John Wiley and Sons, New York, Second Edition.

THOMASIAN A. J. (1963) : A finite criterion for indecomposable channels. *Ann. Math. Stat.* (to appear).

WOLFOWITZ, J. (1961) : *Coding Theorems of Information Theory*, Springer-Verlag, Berlin, and Prentice-Hall, Englewood Cliffs, N. J.

——— (1963) : Products of indecomposable, aperiodic, stochastic matrices, *Proc. Amer. Math. Soc.* (to appear).

Paper received : September, 1962.

SANKHYĀ : THE INDIAN JOURNAL OF STATISTICS

1.	PLACE OF PUBLICATION :	204/1 Barrackpore Trunk Road, Ca lcutt a–5
2.	PERIODICITY OF ITS PUBLICATION :	Quarterly
3.	PRINTER'S NAME :	Kalipada Mukherjee
	NATIONALITY :	Indian
	ADDRESS :	204/1 Barrackpore Trunk Road, Calcutta–35
4.	PUBLISHER'S NAME :	Anikendra Mahalanobis
	NATIONALITY :	Indian
	ADDRESS :	204 Barrackpore Trunk Road, Calcutta–35
5.	EDITORS' NAME :	P. C. Mahalanobis and C. R. Rao
	NATIONALITY :	Indian
	ADDRESS :	203 Barrackpore Trunk Road, Calcutta–35
6.	NAMES AND ADDRESSES OF INDIVIDUALS WHO OWN THE NEWSPAPER AND PARTNERS OR SHAREHOLDERS HOLDING MORE THAN ONE PER CENT OF THE TOTAL CAPITAL :	Statistical Publishing Society (non-profit distributing agency) 204/1 Barrackpore Trunk Road, Calcutta–35

I, Anikendra Mahalanobis, hereby declare that the particulars given above are true to the best of my knowledge and belief.

Sd/ *Anikendra Mahalanobis*

31 July, 1963

BIOMETRIKA

Volume 50, Parts 1 and 2 **June, 1963**

CONTENTS

The subscription, *payable in advance*, is 54/- (or $8.00), per volume (including postage). Cheques should be made payable to *Biometrika*, crossed *"A/c Biometrika Trust"* and sent to the Secretary, Biometrika Office, Department of Statistics, University College, London, W. C. 1. All foreign cheques must be drawn on a Bank having a London agency.

Issued by THE BIOMETRIKA OFFICE, University College, London.

JOURNAL OF THE
ROYAL STATISTICAL SOCIETY

Series B (Methodological)

40*s* each Part. Annual Subscription £3.2*s* (US $9). Post free

CONTENTS OF VOL. 25, PART 1, 1963

Obtainable through any bookseller or direct from the

ROYAL STATISTICAL SOCIETY

21 Bentinck Street, London, W. 1.

THE ANNALS OF MATHEMATICAL STATISTICS

(Founded by H. C. Carver)

The Official Journal of the Institute of Mathematical Statistics

CONTENTS OF VOLUME 34, NO. 2, JUNE, 1963

Subscription Rate: $10.00 per year

Address orders for subscriptions and back numbers to the Secretary, Institute of Mathematical Statistics, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

ECONOMETRICA

Vol. 31, No. 1-2 **CONTENTS** **January-April, 1963**

Subscription price for nonmembers: $12.00 per year in the United States and Canada; $10.00 per year in all other Countries. Single copies $3.00

Graduate Students recommended by a faculty member: $4.00 per year.

Subscriptions should be sent to

The Secretary
ECONOMETRIC SOCIETY
Box 1264, Yale Station, New Haven, Connecticut, U.S.A.

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Editors: P. C. MAHALANOBIS, C. R. RAO

SERIES A, VOL. 25 JULY 1963 PART 2

COMBINATORIAL PROPERTIES OF PARTIALLY BALANCED DESIGNS AND ASSOCIATION SCHEMES*

By R. C. BOSE

University of North Carolina

1. DEFINITION OF ASSOCIATION SCHEMES AND PARTIALLY BALANCED INCOMPLETE BLOCK (PBIB) DESIGNS

Partially balanced designs were first introduced in 1939 by the author in collaboration with Nair (1939) as an extension of the balance incomplete block (BIB) [Yates, 1936b; Fisher and Yates, 1938; Bose, 1939] and lattice (Yates, 1936a) designs. During the twenty four years which have elapsed since then, much theoretical work has been done on the subject and the designs have also been applied to various practical problems. As the initial work was done by both the authors when they were members of the Indian Statistical Institute founded by Professor P. C. Mahalanobis, it seems particularly appropriate to discuss these designs in this volume. The definition of PBIB designs was slightly generalized by Nair and Rao (1942), so as to include as special cases the cubic and other higher dimensional lattices. A further step was taken by Bose and Shimamoto (1952) in introducing the concept of association schemes, and basing the definition PBIB design on these schemes. Since most of the recent work has been on the combinatorial properties of association schemes and PBIB designs based on them, we shall confine ourselves to this topic and leave out the analysis of the designs and methods of constructing them.

Given v treatments, 1, 2, ..., v a relation satisfying the following conditions is said to be an association scheme with m classes :

(a) Any two treatments are either 1st, 2nd, ..., for m-th associates, the relation of association being symmetrical, i.e., if the treatment α is the i-th associate of the treatment β, then β is the i-th associate of the treatment α.

*This research was supported in part by the Mathematics Division of the United States Air Force Office of Scientific Research, under Grant No. AF-AFOSR-84-63.

This paper has been included in *Contributions to Statistics* presented to Professor P. C. Mahalanobis on the occasion of his 70th birthday.

(b) Each treatment has n_i, i-th associates, the number n_i being independent of α.

(c) If any two treatments are i-th associates then the number of treatments which are j-th associates of α and k-th associates of β is p^i_{jk} and is independent of the pair of i-th associates α and β.

The numbers

$$v, n_i, p^i_{jk} \qquad (i, j, k = 1, 2, ..., m), \qquad \ldots \quad (1.1)$$

are the parameters of the association scheme.

If we have an association scheme with m classes, then we get a PBIB design with r replications and b blocks based on the association scheme, if we can arrange the v treatments into b blocks such that

(i) each block contains k treatments (all different),

(ii) each treatment is contained in r blocks,

(iii) if two treatments α and β are i-th associates, then they occur together in λ_i blocks, the number λ_i being independent of the particular pair of i-th associates α and β $(i = 1, 2, ..., m)$.

For a PBIB design based on any association scheme, the parameters of the scheme may be called parameters of the first kind, and the additional parameters

$$b, r, k, \lambda_i \quad (i = 1, 2, ..., m), \qquad \ldots \quad (1.2)$$

may be called parameters of the second kind. Clearly

$$vr = bk, \quad n_1\lambda_1+n_2\lambda_2+\ldots+n_m\lambda_m = r(k-1). \qquad \ldots \quad (1.3)$$

2. Relations among the parameters of association schemes

By definition the number p^i_{jk} is independent of which pair α, β of i-th associates we start with. Consider the pair β, α; we see at once that

$$p^i_{jk} = p^i_{kj}. \qquad \ldots \quad (2.1)$$

The following further relations are easy to prove :

$$\sum_{i=1}^{m} n_i = v-1; \qquad \ldots \quad (2.2)$$

$$\sum_{k=1}^{m} p^i_{jk} = n_j \quad \text{if} \quad i \neq j \qquad \ldots \quad (2.3)$$

$$= n_j-1 \quad \text{if} \quad i = j;$$

$$n_i\, p^i_{jk} = n_j\, p^j_{ik} = n_k\, p^k_{ij}. \qquad \ldots \quad (2.4)$$

These relations were proved by Bose and Nair (1939), in their paper introducing the PBIB designs. These are all the relations in case $m = 2$ but for $m \geqslant 3$ further relations were discovered by Bose and Mesner (1959), and will be discussed in a subsequent section.

It is useful to make a convention that *each treatment is its own zeroth associate and of no other treatments.* Then clearly we must take,

$$n_0 = 1; \qquad \dots \quad (2.5)$$

$$\begin{aligned} p^0_{ji} = p^0_{ij} &= 0 \qquad \text{if } i \neq j, \\ &= n_j \qquad \text{if } i = j; \end{aligned} \qquad \dots \quad (2.6)$$

$$\begin{aligned} p^i_{k0} = p^i_{0k} &= 0 \qquad \text{if } i \neq k, \\ &= 1 \qquad \text{if } i = k. \end{aligned} \qquad \dots \quad (2.7)$$

We can now write (2.2) and (2.3) as

$$\sum_{i=0}^{m} n_i = v \qquad \dots \quad (2.8)$$

$$\sum_{k=0}^{m} p^i_{jk} = n_j \qquad \dots \quad (2.9)$$

for $i, j, k = 0, 1, \dots m$. It should also be noted that (2.4) remains valid if one or more of i, k, j is zero.

Also for a PBIB design based on the association scheme we must have

$$\sum_{i=0}^{m} n_i \lambda_i = rk, \qquad \text{where } \lambda_0 = r. \qquad \dots \quad (2.10)$$

3. A less demanding definition of two class association schemes

The definition given in Section 1, for association schemes is not minimal, i.e. the constancy of some of the parameters can be deduced from the others. For two class association schemes Bose and Clatworthy (1955) proved the following lemma :

Lemma 1 : *Let there exist a relationship of association among v treatments satisfying the conditions :* (a) *Any two treatments are either first associates or second associates,* (b) *each treatment has n_1 first associates and n_2 second associates,* (c) *for any pair of treatments which are first associates the number p^1_{11} of treatments common to the first associates of the first, and the first associates of the second is independent of the pair of treatments with which we start.*

Then, for every pair of first associates among the v treatments, the members p^1_{12}, p^1_{21} and p^1_{22} are constants and $p^1_{12} = p^1_{21}$.

Lemma 2 : *Let there exist a relationship of association among v treatments satisfying conditions* (a) *and* (b) *of Lemma* 1, *and the condition* (c) *for any pair of treatments which are second associate, the number p^2_{11} of treatments which are common to the first associates of the first, and the first associates of the second is independent of the pair of treatments with which we start.*

Then, for every pair of second associates among the v treatments, the numbers p^2_{12}, p^2_{21} and p^2_{22} are constants and $p^2_{12} = p^2_{21}$.

One can ask whether one of the preceding lemmas implies the other. The answer is no. Consider the association scheme with $v = 7$, for which the first and second associates are shown below :

treatment	first associates	second associates
1	2, 4, 5, 7	3, 6
2	3, 5, 6, 1	4, 7
3	4, 6, 7, 2	5, 1
4	5, 7, 1, 3	6, 2
5	6, 1, 2, 4	7, 3
6	7, 2, 3, 5	1, 4
7	1, 3, 4, 6	2, 5

Here $n_1 = 4, n_2 = 2$. For any pair of treatments α, β which are second associates p^2_{11} is 3 independently of α, β. Hence Lemma 2 is satisfied with $p^2_{12} = p^2_{21} = 1$, $p^2_{22} = 0$. However, for any two treatments α and β which are first associates p^1_{11} is either 1 or 2.

In view of the Lemmas 1 and 2, the condition (c) in the definition of m class association schemes given in Section 1, may be replaced in the special case $m = 2$, by the condition (c′), for any pair of treatments which are i-th associates the number p^i_{11} for $i = 1, 2$, of the treatments which are common to the first associates of first and first associates of the second is independent of the pair of treatments with which we start.

For a two class association scheme, the values of the parameters p^i_{jk} $(i, j, k = 1, 2)$, may conveniently be written in the form of two symmetric matrices

$$P_1 = (p^1_{jk}) = \begin{pmatrix} p^1_{11} & p^1_{12} \\ p^1_{21} & p^1_{22} \end{pmatrix}, \quad P_2 = (p^2_{jk}) = \begin{pmatrix} p^2_{11} & p^2_{12} \\ p^2_{21} & p^2_{22} \end{pmatrix}. \qquad \ldots \quad (3.1)$$

4. Some examples of two class association schemes and PBIB designs based on them

We shall give below some simple examples of two class association schemes, and a few designs based on them. This enumeration is for illustrative purposes and is not exhaustive.

(a) *The group divisible (GD) association scheme.* In this case there are mn treatments, which are divided into m groups of n treatments each. Two treatments belonging to the same group are first associates, and two treatments belonging to different groups are second associates. The association scheme can be exhibited by writing down the mn treatments in the form of a rectangular array, the treatments of the same group occupying the same column. It is readily seen that the parameters of the association scheme so obtained are

$$v = mn, \; n_1 = n-1, \; n_2 = n(m-1), \qquad \ldots \quad (4.1)$$

$$P_1 = \begin{pmatrix} n-2 & 0 \\ 0 & n(m-1) \end{pmatrix}, \quad P_2 = \begin{pmatrix} 0 & n-1 \\ n-1 & n(m-2) \end{pmatrix}. \qquad \ldots \quad (4.2)$$

For example let $m = 4$, $n = 3$. The corresponding GD association scheme is

$$\begin{matrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12. \end{matrix} \qquad \dots \quad (4.3)$$

The first associates of the treatment 1 are 5 and 9, and the second associates are 2, 3, 4, 6, 7, 8, 10, 11, 12.

A PBIB design based on the above association scheme for which the parameters of the second kind are

$$b = 9, \quad r = 3, \quad k = 4, \quad \lambda_1 = 1, \quad \lambda_2 = 0, \qquad \dots \quad (4.4)$$

is given below :

$$\begin{matrix} (1, & 8, & 2, & 3), & (8, & 5, & 6, & 7), & (5, & 2, & 11, & 4), \\ (6, & 11, & 12, & 1), & (11, & 10, & 9, & 8), & (10, & 12, & 3, & 5), \\ (12, & 9, & 7, & 2), & (7, & 4, & 1, & 10), & (9, & 3, & 4, & 6). \end{matrix} \qquad \dots \quad (4.5)$$

(b) *The triangular association scheme.* We take an $m \times m$ square, and fill-in the $m(m-1)/2$ positions above the leading diagonal by different treatments, taken in any order. The positions in the leading diagonal are left blank, while positions below this diagonal are filled so that the scheme is symmetrical with respect to the diagonal. Two treatments in the same row (or same column) are first associates. Two treatments which do not occur in the same row or same column are second associates. It is readily verified that the parameters of the association scheme so obtained are

$$v = m(m-1)/2, \; n_1 = 2m-4, \quad n_2 = (m-2)(m-3)/2, \qquad \dots \quad (4.6)$$

$$P_1 = \begin{pmatrix} m-2 & m-3 \\ m-3 & (m-3)(m-4)/2 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 4 & 2m-8 \\ 2m-8 & (m-4)(m-5)/2 \end{pmatrix}. \qquad \dots \quad (4.7)$$

This scheme is called the triangular association scheme.

An interesting class of PBIB designs based on the triangular association scheme is formed by taking for blocks, treatments in the same row of the association scheme. For these designs we will have

$$b = m, \quad r = 2, \quad k = m-1, \quad \lambda_1 = 1, \quad \lambda_2 = 0.$$

As an illustration let us take $m = 5$. The association scheme is then

$$\begin{matrix} * & 1 & 2 & 3 & 4 \\ 1 & * & 5 & 6 & 7 \\ 2 & 5 & * & 8 & 9 \\ 3 & 6 & 8 & * & 10 \\ 4 & 7 & 9 & 10 & * \end{matrix} \qquad \dots \quad (4.8)$$

where two different treatments are first associates if they occur together in the same row (or same column) of the above scheme, and second associates otherwise. The parameters of the association scheme are

$$v = 10, \quad n_1 = 6, \quad n_2 = 3. \qquad \dots \quad (4.9)$$

$$P_1 = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 4 & 2 \\ 2 & 0 \end{pmatrix}. \qquad \dots \quad (4.10)$$

By taking the rows of (4.8) for blocks we get the design with parameters of the second kind

$$b = 5, \quad r = 2, \quad k = 4, \quad \lambda_1 = 1, \quad \lambda_2 = 0. \qquad \dots \quad (4.11)$$

There may be more than one PBIB designs based on the same association scheme. For example there exist at least three other designs, besides (4.11), based on the triangular association scheme (4.8). We give below the parameters of the second kind, (4.12), (4.14) and (4.16) together with the corresponding blocks (4.13), (4.15), (4.17) for these designs :

$$b = 10, \quad r = 3, \quad k = 3, \quad \lambda_1 = 1, \quad \lambda_2 = 0. \qquad \dots \quad (4.12)$$

$$\begin{array}{ccccc} (1,\ 2,\ 5), & (8,\ 9,\ 10), & (2,\ 8,\ 3), & (7,\ 5,\ 9), & (9,\ 4,\ 2), \\ (5,\ 6,\ 8), & (3,\ 10,\ 4), & (10,\ 7,\ 6), & (4,\ 1,\ 7), & (6,\ 3,\ 2). \end{array} \qquad \dots \quad (4.13)$$

$$b = 6, \quad r = 3, \quad k = 5, \quad \lambda_1 = 1, \quad \lambda_2 = 2. \qquad \dots \quad (4.14)$$

$$\begin{array}{ccc} (1,\ 8,\ 9,\ 7,\ 3), & (1,\ 8,\ 4,\ 10,\ 5), & (8,\ 4,\ 6,\ 7,\ 2), \\ (4,\ 6,\ 9,\ 5,\ 3), & (1,\ 6,\ 9,\ 10,\ 2), & (10,\ 7,\ 5,\ 2,\ 3). \end{array} \qquad \dots \quad (4.15)$$

$$b = 10, \quad r = 4, \quad k = 4, \quad \lambda_1 = 1, \quad \lambda_2 = 2. \qquad \dots \quad (4.16)$$

$$\begin{array}{ccccc} (2,\ 10,\ 6,\ 7), & (10,\ 1,\ 2,\ 5), & (7,\ 3,\ 8,\ 2), & (6,\ 2,\ 9,\ 4), & (1,\ 9,\ 10,\ 8), \\ (5,\ 4,\ 3,\ 10), & (8,\ 7,\ 4,\ 1), & (3,\ 5,\ 7,\ 9), & (9,\ 6,\ 1,\ 3), & (4,\ 8,\ 5,\ 6). \end{array} \qquad \dots \quad (4.17)$$

(c) *The singly linked block (SLB) association scheme.* Consider a balanced incomplete block (BIB) design D with b treatments, v blocks, k replications, block size r and $\lambda = 1$, i.e. every pair of treatments occurs in exactly one block. Then

$$bk = vr, \qquad b-1 = k(r-1). \qquad \dots \quad (4.18)$$

Consider v new treatments each corresponding to one block of D. Two of these new treatments will be called first associates if the corresponding blocks of D have a common treatment and second associates if the corresponding blocks of D have no common treatment. Shrikhande (1950) has shown that this association relation satisfies the conditions (a), (b), (c) of Section 1 with parameters,

$$v = k(kr-k+1)/r, \quad n_1 = r(k-1), \quad n_2 = (k-r)(r-1)(k-1)/r, \qquad \dots \quad (4.19)$$

$$P_1 = \begin{pmatrix} (k-2)+(r-1)^2 & (r-1)(k-r) \\ (r-1)(k-r) & (r-1)(k-r)(k-r-1)/r \end{pmatrix}, \qquad \dots \quad (4.20)$$

$$P_2 = \begin{pmatrix} r^2 & r(k-r-1) \\ r(k-r-1) & \dfrac{(k-r)(r-1)(k-1)}{r} - r(k-r-1)-1 \end{pmatrix}. \qquad \dots \quad (4.21)$$

This association scheme is defined to be an SLB scheme. Every BIB design with $\lambda = 1$ gives rise to such a scheme.

Let D^* be the dual of the design D, i.e. the blocks and treatments of D^* correspond to the treatments and blocks of D, and a block of D^* contains a treatment of D^* if and only if the corresponding treatment of D, occurs in the corresponding block of D. Then D^* is a PBIB design based on the association scheme just described, for which the parameters of the second kind are

$$b, \quad r, \quad k, \quad \lambda_1 = 1, \quad \lambda_2 = 0. \qquad \ldots \quad (4.22)$$

These designs are known as singly linked block designs.

In the special case $r = 2$, when the BIB design D consists of all possible pairs of r treatments, the SLB scheme reduces to a triangular scheme with $m = k+1$.

(d) *The Latin square* (L_r) *association scheme.* Consider $v = k^2$ treatments which may be set forth in a $k \times k$ scheme. Thus if $k = 4$ and the treatments are 1, 2, ..., 16, we have the scheme

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

... (4.23)

For the case $r = 2$, we define two treatments as first associates if they occur in the same row or column of the square scheme, and second associates otherwise. The association scheme so defined may be called the L_2 association scheme. The parameters of the L_2 scheme are

$$v = k^2, \quad n_1 = 2(k-1), \qquad n_2 = (k-1)^2, \qquad \ldots \quad (4.24)$$

$$P_1 = \begin{pmatrix} k-2 & k-1 \\ k-1 & (k-1)(k-2) \end{pmatrix}, \quad P_2 = \begin{pmatrix} 2 & 2(k-2) \\ 2(k-2) & (k-2)^2 \end{pmatrix}. \qquad \ldots \quad (4.25)$$

In the general case $2 \leqslant r \leqslant k+1$, we take a set of $r-2$ mutually orthogonal Latin squares (if such a set exists). For an L_r association scheme we then define two treatments to be first associates if they occur together in the same row or column of the square scheme, or if they correspond to the same symbol of one of the Latin squares. Otherwise we define them to be second associates. For example if $k = 4$, $r = 4$ and we take the Latin squares

$[L_1]$

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

$[L_2]$

1	2	3	4
3	4	1	2
4	3	2	1
2	1	4	3

... (4.26)

then the first associates of the treatment 7 are 5, 6, 8, 3, 11, 15, 4, 10, 13, 1, 12, 14 because the treatment 7 corresponds to the symbol 4 in $[L_1]$ and the symbol 1 in $[L_2]$. The parameters of the L_r association scheme are given by

$$v = k^2,\ n_1 = r(k-1),\quad n_2 = (k-1)(k-r+1), \qquad \dots \quad (4.27)$$

$$P_1 = \begin{pmatrix} (k-2)+(r-1)(r-2) & (r-1)(k-r+1) \\ (r-1)(k-r+1) & (k-r)(k-r+1) \end{pmatrix}, \qquad \dots \quad (4.28)$$

$$P_2 = \begin{pmatrix} r(r-1) & r(k-r) \\ r(k-r) & (k-r)^2+(r-2) \end{pmatrix}. \qquad \dots \quad (4.29)$$

The simplest class of PBIB designs based on the L_r schemes are the lattice designs, first introduced by Yates (1936a), which are obtained as follows :

We obtain k blocks each of size k by putting in the same block all treatments occurring in the same row of the square scheme. This gives us one complete replication. Another complete replication is obtained in the same way from columns. Again from each of the $r-2$ mutually orthogonal Latin squares we obtain k blocks giving a complete replication by putting all those treatments which correspond to the same symbol, in the same block. The parameters of the second kind for the Lattice design so obtained are

$$b = kr,\ \ r,\ k,\ \lambda_1 = 1,\quad \lambda_2 = 0. \qquad \dots \quad (4.30)$$

For example for the case $k = 4$, $r = 4$, the lattice design based on the square scheme (4.23) and the Latin squares $[L_1]$ and $[L_2]$ in (4.26), is

Rep 1,	(1, 2, 3, 4),	(5, 6, 7, 8),	(9, 10, 11, 12),	(13, 14, 15, 16),
Rep 2,	(1, 5, 9, 13),	(2, 6, 10, 14),	(3, 7, 11, 15),	(4, 8, 12, 16),
Rep 3,	(1, 6, 11, 16),	(2, 5, 12, 15),	(3, 8, 9, 14),	(4, 7, 10, 13),
Rep 4,	(1, 7, 12, 14),	(2, 8, 11, 13),	(3, 5, 10, 16),	(4, 6, 9, 15),

... (4.31)

If we had chosen only the Latin square $[L_1]$, the design obtained would have been given by the first 3 rows of (4.31). Let us call this design D_3. To illustrate the fact that there may be designs other than lattices based on an L_r association scheme, we give below another design based on the same L_3 scheme as D_3. The parameters of the second kind for this design are

$$b = 16,\ \ r = 3,\quad k = 3,\quad \lambda_1 = 0,\quad \lambda_2 = 1, \qquad \dots \quad (4.32)$$

and the blocks are

(10, 7, 16),	(11, 8, 13),	(12, 5, 14),	(9, 6, 15),
(13, 4, 6),	(14, 1, 7),	(15, 2, 8),	(16, 3, 5),
(7, 9, 4),	(8, 10, 1),	(5, 11, 2),	(6, 12, 3),
(2, 16, 9),	(3, 13, 10),	(4, 14, 11),	(1, 15, 12).

... (4.33)

(e) *The cyclic association scheme.* Let the v treatments be the v integers denoted by 1, 2, ..., v. Let

$$d_1, d_2, \ldots, d_{n_1} \qquad \ldots \quad (4.34)$$

be a set of n_1 integers satisfying the conditions

(i) the d's are all different, and $0 < d_j < v$ ($j = 1, 2, \ldots, n_1$),

(ii) among the $n_1(n_1-1)$ differences $d_j - d_{j'}$, ($j, j' = 1, 2, \ldots, n_1, j \neq j'$) reduced (mod v), each of the numbers $d_1, d_2, \ldots, d_{n_1}$ occurs g times, whereas each of the numbers $e_1, e_2, \ldots, e_{n_2}$ occurs h times, where $d_1, d_2, \ldots, d_{n_1}, e_1, e_2, \ldots, e_{n_2}$ are all the different $v-1$ numbers $1, 2, \ldots, v-1$.

Clearly it is necessary that

$$n_1 g + n_2 h = n_1(n_1-1). \qquad \ldots \quad (4.35)$$

Let the first associates of the treatment i be

$$i+d_1, i+d_2, \ldots, i+d_{n_1} \pmod v, \qquad \ldots \quad (4.36)$$

and the remaining treatments be second associates of i. The association thus defined is a two class association scheme with parameters v, n_1, n_2,

$$P_1 = \begin{pmatrix} g & n_1-g-1 \\ n_1-g-1 & n_2-n_1+g+1 \end{pmatrix}, \qquad \ldots \quad (4.37)$$

$$P_2 = \begin{pmatrix} h & n_1-h \\ n_1-h & n_2-n_1+h-1 \end{pmatrix}. \qquad \ldots \quad (4.38)$$

For example taking $v = 13$, and

$$d_1 = 2, \quad d_2 = 5, \quad d_3 = 6, \quad d_4 = 7, \quad d_5 = 8, \quad d_6 = 11$$

we find that among the differences $d_j - d_{j'}$ reduced (mod 13) each d occurs twice, whereas every other non-zero integer less than 13 occurs thrice. Hence the conditions (i), (ii) are satisfied. We thus have a cyclic association scheme with parameters

$$v = 13, \quad n_1 = 6, \quad n_2 = 6, \qquad \ldots \quad (4.39)$$

$$P_1 = \begin{pmatrix} 3 & 2 \\ 3 & 3 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 3 & 3 \\ 3 & 2 \end{pmatrix}. \qquad \ldots \quad (4.40)$$

An example of a PBIB design based on this association scheme, is provided by the design with the second kind of parameters

$$b = 13, \quad r = 3, \quad k = 3, \quad \lambda_1 = 1, \quad \lambda_2 = 0, \qquad \ldots \quad (4.41)$$

and the blocks

$$(1, 3, 9), \quad (2, 4, 10), \quad (3, 5, 11), \quad (4, 6, 12), \quad (5, 7, 13),$$
$$(6, 8, 1), \quad (7, 9, 2), \quad (8, 10, 3), \quad (9, 11, 4), \quad (10, 12, 5),$$
$$(11, 13, 6), \quad (12, 1, 7), \quad (13, 2, 8). \qquad \ldots \quad (4.42)$$

5. Association matrices and the algebra of association schemes

Consider an m class association scheme. The i-th association matrix is defined by

$$B_i = (b^{\beta}_{\alpha i}) = \begin{pmatrix} b^1_{1i} & b^2_{1i} & \ldots & b^v_{1i} \\ \ldots & \ldots & \ldots & \ldots \\ b^1_{vi} & b^2_{vi} & \ldots & b^v_{vi} \end{pmatrix} \qquad \ldots \quad (5.1)$$

where $\qquad b^{\beta}_{\alpha i} = 1$ if the treatments α and β are i-th associates

$$= 0 \quad \text{otherwise.} \qquad \ldots \quad (5.2)$$

Hence B_i is a symmetric $v \times v$ matrix in which the element in the α-th row and β-th column is unity if the treatments α and β are i-th associates and 0 otherwise. The total of each row and each column in B_i is n_i. Clearly

$$B_0 = I_v, \quad \text{the } v \times v \text{ identity matrix.} \qquad \ldots \quad (5.3)$$

It is also readily seen that

$$B_0 + B_1 \ldots + B_m = J_v \qquad \ldots \quad (5.4)$$

where J_v is the $v \times v$ matrix each of whose elements is unity. It is also clear that the linear form

$$c_0 B_0 + c_1 B_1 + \cdots + c_m B_m \qquad \ldots \quad (5.5)$$

is equal to the zero matrix if and only if

$$c_0 = c_1 = \cdots = c_m = 0.$$

Hence linear functions of $B_0, B_1, \ldots, B_m$ form a vector space with basis $B_0, B_1, \ldots, B_m$.

Thompson (1954, 1958) and independently Mesner (1956) proved the fundamental formula

$$B_j B_k = p^0_{jk}\, B_0 + p^1_{jk}\, B_1 + \cdots + p^m_{jk}\, B_m. \qquad \ldots \quad (5.6)$$

This shows that the product of any two matrices of the form (5.5), may be expressed as a linear combination of terms of the form $B_j B_k$ and will reduce to the form (5.5). The set of matrices of this form is closed under multiplication. We will here confine ourselves to the case when the coefficients c_i range over the field of real numbers.

The linear functions of the association matrices $B_0, B_1, \ldots, B_m$ form a linear associative algebra. Using the fact

$$B_i(B_jB_k) = (B_iB_j)B_k,$$

one can show that (Bose and Mesner, 1959),

$$\sum_u p_{ij}^u \, p_{uk}^t = \sum_u p_{jk}^u \, p_{iu}^t. \qquad \ldots \quad (5.7)$$

where u runs from 0 to m and the remaining indices are arbitrary but fixed and

$$0 \leqslant i, j, k, t \leqslant m.$$

Now let us define Π_k by

$$\Pi_k = (p_{ik}^j) = \begin{pmatrix} p_{0k}^0 & p_{0k}^1 & \cdots & p_{0k}^m \\ p_{ik}^0 & p_{ik}^1 & \cdots & p_{ik}^m \\ \cdots & \cdots & \cdots & \cdots \\ p_{mk}^0 & p_{mk}^1 & \cdots & p_{mk}^m \end{pmatrix}, \quad k = 0, 1, \ldots, m. \qquad \ldots \quad (5.8)$$

Then (5.7) is equivalent to

$$\Pi_j \; \Pi_k = p_{jk}^0\Pi_0 + p_{jk}^1\Pi_1 + \cdots + p_{jk}^m\Pi_m. \qquad \ldots \quad (5.9)$$

Thus the matrices Π_k given by (5.7) multiply in the same manner as the association matrices $B_k (k = 0, 1, \ldots, m)$. From (2.7) it follows that if

$$c_0\Pi_0 + c_1\Pi_1 + \cdots + c_m\Pi_m = 0,$$

then

$$c_0 = c_1 = \cdots = c_m = 0,$$

i.e. $\Pi_0, \Pi_1, \ldots, \Pi_m$ are linearly independent. They thus form the basis for a vector space and combine in the same way as the B's under addition as well as multiplication. They provide a regular representation in $(m+1)\times(m+1)$ matrices of the algebra given by the B's, which are $v \times v$ martices. In particular

$$\Pi_0 = I_{m+1}. \qquad \ldots \quad (5.10)$$

This regular representation was discovered independently by Bose (1955) and by Mesner (1956), and further properties were studied in a joint paper (1959) by both. Some of these will be mentioned later.

Since the B's are commutative, the Π's are commutative. In general they are not incidence matrices and are not symmetric. In analogy with (5.4), all the elements of the row j of $\sum_k \Pi_k$ are equal to n_j. Let

$$B = c_0B_0 + c_1B_1 + \cdots + c_mB_m, \qquad \ldots \quad (5.11)$$

be any element of our algebra, and let $f(\lambda)$ be a polynomial. Then we can express

$$f(B) = l_0B_0 + l_1B_1 + \cdots + l_mB_m. \qquad \ldots \quad (5.12)$$

If

$$\Pi = c_0\Pi_0 + c_1\Pi_1 + \cdots + c_m\Pi_m, \qquad \ldots \quad (5.13)$$

is the representation of B, then

$$f(\Pi) = l_0\Pi_0+l_1\Pi_1+\cdots+l_m\Pi_m. \quad \ldots \quad (5.14)$$

Let $f(\lambda)$ be the minimum function of B, i.e. $f(\lambda)$ is the monic polynomial of least degree for which

$$f(B) = 0.$$

Similarly let $\phi(\lambda)$ be the minimum function of Π.

Then $\qquad f(B) = 0 \rightarrow l_0 = l_1 = \cdots = l_m = 0 \rightarrow f(\Pi) = 0$

i.e. $f(\lambda)$ is divisible by $\phi(\lambda)$. In the same way $\phi(\lambda)$ is divisible by $f(\lambda)$. Since both are monic polynomials

$$f(\lambda) = \phi(\lambda).$$

Hence B and Π have the same distinct characteristic roots, and every matrix B, has utmost $m+1$ distinct roots, which are solutions of the minimum equation of Π. In general the number of treatments is much larger than the number of classes m, so that the roots of B have high multiplicities.

Since $B_jB_k = B_kB_j$ it follows that

$$\Pi_j\,\Pi_k = \Pi_k\,\Pi_j. \quad \ldots \quad (5.15)$$

It is easy to show that the relations (2.4), (2.8), (2.9) among the parameters of the association scheme are consequences of (5.15). However, this relation leads to new identities when $m > 2$. For example in the case $m = 3$, Bose and Mesner (1959) show that there is a new identity

$$x^2\left(\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}\right)+x\left(\frac{a_{12}}{n_1}+\frac{a_{23}}{n_2}+\frac{a_{31}}{n_3}-n_1-n_2-n_3-1\right)$$

$$+\left(\frac{a_{31}\,a_{12}}{n_1}+\frac{a_{12}\,a_{23}}{n_2}+\frac{a_{23}\,a_{31}}{n_3}-n_1a_{23}-n_2a_{31}-n_3a_{12}+n_1n_2n_3\right) = 0, \quad \ldots \quad (5.16)$$

where $\qquad a_{12} = n_1p^1_{22} = n_2p^2_{12}, \quad a_{31} = n_3p^3_{11} = n_1p^1_{13}, \quad a_{23} = n_2p^2_{23} = n_3p^3_{23}$

$$x = n_1p^1_{23} = n_2p^2_{13} = n_3p^3_{12}.$$

This identity is not derivable from (2.4), (2.8), (2.9).

Given a set of positive integral parameters v, n_j, p^i_{jk} $(i, j, k = 0, 1, \ldots, m)$, satisfying (2.5), (2.6), (2.7), one may ask whether it is possible for an m class association scheme with these parameters to exist. Then the matrix equations (5.15) provide necessary condition for existence.

6. Combinatorial applications of the algebra of association matrices

(a) Consider a PBIB design based on an m class association scheme, with the association matrices B_i defined by (5.1). Let

$$N = (n_{ij}), \quad i = 1, 2, \ldots v, \quad j = 1, 2, \ldots, b\,; \qquad \ldots \quad (6.1)$$

be the incidence matrix of the design, i.e. $n_{ij} = 1$ or 0 according as the treatment i does or does not occur in the j-th block. Then

$$B = NN' = rB_0 + \lambda_1 B_1 + \cdots + \lambda_m B_m, \qquad \ldots \quad (6.2)$$

$$\Pi = r\Pi_0 + \lambda_1 \Pi_1 + \cdots + \lambda_m \Pi_m. \qquad \ldots \quad (6.3)$$

The elements of NN' are non-negative and for connected designs, i.e. designs in which every treatment contrast is estimable, NN' is irreducible. Also in virtue of the identity (2.10) the sum of the elements in every row of NN' is rk. Hence

$$B^* = \frac{1}{rk}\, B = \frac{1}{rk}\, NN', \qquad \ldots \quad (6.4)$$

is a stochastic matrix (i.e. an irreducible matrix for which the sum of each row is unity). For such a matrix (Brauer, 1952), unity is a simple root and is greater than all the other roots. Hence rk is a simple root of B and is, therefore, also a simple root of Π. Let

$$p_{ij} = rp^j_{i0} + \lambda_1 p^j_{i1} + \cdots + \lambda_m p^j_{im}. \qquad \ldots \quad (6.5)$$

Then

$$\Pi = (p_{ij}).$$

If θ is a characteristic root of Π we have

$$\begin{vmatrix} p_{00}-\theta & p_{01} & \cdots & p_0 \\ p_{10} & p_{11}-\theta & \cdots & p_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ p_{m0} & p_{m1} & \cdots & p_{mm}-\theta \end{vmatrix} = 0.$$

Now from (2.5)—(2.10),

$$\sum_{i=0}^{m} p_{ij} = r \sum_{i=0}^{m} p^j_{i0} + \lambda_1 \sum_{i=0}^{m} p^j_{i1} + \cdots + \lambda_m \sum_{i=0}^{m} p^j_{im}$$

$$= n_0 r + n_1\lambda_1 +, \ldots + n_m\lambda_m = rk.$$

Hence

$$(rk-\theta)\begin{vmatrix} 1 & 1 & \cdots & 1 \\ p_{10} & p_{11}-\theta & \cdots & p_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ p_{m0} & p_{m1} & \cdots & p_{mm} \end{vmatrix} = 0.$$

Therefore the characteristic roots of Π other than rk, are roots of the matrix

$$\Pi^* = (p^*_{ij}) \qquad i, j = 1, 2, \ldots, m; \qquad \ldots \quad (6.6)$$

where

$$\begin{aligned} p^*_{ij} &= p_{ij} - p_{i0} \\ &= r\delta_{ij} + \lambda_1 p^i_{i1} + \ldots + \lambda_m p^j_{im} - n_i \lambda_i \end{aligned} \qquad \ldots \quad (6.7)$$

δ_{ij} being the Kronecker delta.

(b) Fisher (1940) showed that for a balanced incomplete block (BIB) design there holds the inequality

$$b \geqslant v \qquad \ldots \quad (6.8)$$

where b is the number of blocks and v is the number of treatments. (For an alternative proof see Bose (1949)). One may ask what the corresponding result is for PBIB designs. Now

$$\begin{aligned} b &\geqslant \text{rank } N \\ &\geqslant \text{rank } NN'. \end{aligned}$$

But the rank of NN' is v, unless NN' is singular, i.e., has a zero characteristic root, in which case Π and therefore Π^* has a zero characteristic root. Thus:

A necessary condition for $b \leqslant v$ in a PBIB design is

$$|p^*_{ij}| = 0 \qquad \ldots \quad (6.9)$$

where p^*_{ij} is given by (6.7).

Thus Fisher's inequality $b \geqslant v$ is satisfied in general. It can be violated by only those designs for which (6.9) is satisfied. This result is due to Nair (1943). An alternative proof will be found in Bose (1952).

(c) The multiplicities of the characteristic roots of $B = NN'$ can be calculated. Let $\theta_0 = rk$, $\theta_1, \ldots, \theta_m$ be these characteristic roots with multiplicities $\alpha_0 = 1$, α_1, $\alpha_2, \ldots, \alpha_m$. Then $\theta_0, \theta_1, \ldots, \theta_m$ are the characteristic roots of Π. Remembering that the trace of any matrix is equal to the sum of its characteristic roots we have

$$\alpha_0\, \theta_0^n + \alpha_1\, \theta_1^n + \cdots + \alpha_m \theta_m^n = \text{tr } B^n, \quad n = 0, 1, 2, \ldots . \qquad \ldots \quad (6.10)$$

Using the fundamental formula (5.6), we may express B^n in the form

$$B^n = c_{0n} B_0 + c_{1n} B_1 +, \ldots + c_{mn} B_m. \qquad \ldots \quad (6.11)$$

Then since B_0 is the only B with non-zero diagonal elements

$$\text{tr } B^n = v c_{0n}. \qquad \ldots \quad (6.12)$$

We shall illustrate the determination of the multiplicities α_i by considering the case $m = 2$. In this case θ_1 and θ_2 are the roots of the matrix

$$\Pi^* = \begin{pmatrix} r+\lambda_1 p^1_{11}+\lambda_2 p^1_{12}-n_1\lambda_1 & \lambda_1 p^2_{11}+\lambda_2 p^2_{12}-n_1\lambda_1 \\ \lambda_1 p^1_{21}+\lambda_2 p^1_{22}-n_2\lambda_2 & r+\lambda_1 p^2_{21}+\lambda_2 p^2_{22}-n_2\lambda_2 \end{pmatrix}. \quad \ldots \quad (6.13)$$

Using the identities connecting the parameters we get after some computation the result that θ_1, θ_2 satisfy the quadratic

$$(r-\theta)^2+[(\lambda_1-\lambda_2)(p^2_{12}-p^1_{12})-(\lambda_1+\lambda_2)](r-\theta)$$
$$+[(\lambda_1-\lambda_2)(\lambda_2 p^1_{12}-\lambda_1 p^2_{12})+\lambda_1\lambda_2] = 0. \quad \ldots \quad (6.14)$$

Setting $$\gamma = p^2_{12}-p^1_{12}, \quad \beta = p^1_{12}+p^2_{12}, \quad \Delta = \gamma^2+2\beta+1, \quad \ldots \quad (6.15)$$

we have $$\theta_1 = r+\tfrac{1}{2}[(\lambda_1-\lambda_2)(\sqrt{\Delta}+\gamma)-(\lambda_1+\lambda_2)], \quad \ldots \quad (6.16)$$

$$\theta_2 = r-\tfrac{1}{2}[(\lambda_1-\lambda_2)(\sqrt{\Delta}-\gamma)+(\lambda_1+\lambda_2)].$$

Now from (6.10) $$\text{tr } I = 1+\alpha_1+\alpha_2 = v, \quad \ldots \quad (6.17)$$

$$\text{tr } NN' = rk+\alpha_1\theta_1+\alpha_2\theta_2 = vr, \quad \ldots \quad (6.18)$$

whence after some calculation

$$\alpha_1 = \frac{n_1+n_2}{2}-\frac{(n_1-n_2)+\gamma(n_1+n_2)}{2\sqrt{\Delta}} \quad \ldots \quad (6.19)$$

$$\alpha_2 = \frac{n_1+n_2}{2}+\frac{(n_1-n_2)+\gamma(n_1+n_2)}{2\sqrt{\Delta}}. \quad \ldots \quad (6.20)$$

It is interesting to note that the multiplicities α_1 and α_2 depend only on the parameters of the association scheme, i.e. parameters of the first kind. This is a special case of the following result due to Bose and Mesner (1954) :

If $B = c_0B_0+c_1B_1+\cdots+c_mB_m$ is a linear function of the association matrices of an association scheme, then

$$|B-I\theta| = (\theta_0-\theta)^{\alpha_0}(\theta_1-\theta)^{\alpha_1}\ldots(\theta_m-\theta)^{\alpha_m}, \quad \ldots \quad (6.21)$$

where the multiplicities $\alpha_0, \alpha_1, \ldots, \alpha_m$ are independent of $c_0, c_1, \ldots, c_m$. This remains true if B is given by (6.2). Hence the multiplicity of the roots of NN' is independent of $r, \lambda_1, \lambda_2, \ldots, \lambda_m$.

Since the multiplicities α_i are expressible in terms of the parameters of the association scheme, we cannot have a set of parameters leading to nonintegral values α_i. This fact can be used to prove the impossibility of certain association schemes.

We are now in a position to see how Fisher's inequality should be modified for the case when Π and therefore Π^* has a zero characteristic root. Let α be the multiplicity of the root zero of $B = NN'$. Then

$$\begin{aligned} b &\geqslant \text{rank } N \\ &\geqslant \text{rank } NN' \\ &= v-\alpha. \end{aligned}$$

Hence Fisher's inequality is replaced by

$$b \geqslant v-\alpha. \qquad \ldots \quad (6.22)$$

This result is due to Connor and Clatworthy (1954).

(d) Since the distinct roots $\theta_0, \theta_1, \ldots, \theta_m$ of NN' are the roots of Π, the θ's are functions of parameters of the design. Since NN' is symmetric the θ's are all real and positive. This fact yields interesting inequalities among the parameters of the design.

For PBIB designs based on two class association schemes we get the inequalities,

$$r \geqslant \tfrac{1}{2}[(\lambda_2-\lambda_1)(\sqrt{\Delta}+\gamma)+(\lambda_1+\lambda_2)], \qquad \ldots \quad (6.23)$$

$$r \geqslant \tfrac{1}{2}[(\lambda_1-\lambda_2)(\sqrt{\Delta}-\gamma)+(\lambda_1+\lambda_2)]. \qquad \ldots \quad (6.24)$$

(e) We shall illustrate some of the above results with reference to group divisible PBIB designs, based on the group divisible (GD) association scheme, defined earlier. From (4.2) and (6.15) we have

$$\gamma = n-1, \quad \beta = n-1, \qquad \Delta = n^2. \qquad \ldots \quad (6.25)$$

Hence from (6.16), (6.19), (6.20), using the identities connecting the parameters

$$\theta_1 = r+n(\lambda_1-\lambda_2)-\lambda_1 = rk-v\lambda_2, \quad \theta_2 = r-\lambda_1. \qquad \ldots \quad (6.26)$$

$$\alpha_1 = m-1, \quad \alpha_2 = v-m. \qquad \ldots \quad (6.27)$$

We therefore have the following result:

Group divisible designs can be divided into three classes :

(i) *Regular group divisible designs for which $rk-v\lambda_2 > 0$, $r-\lambda_1 > 0$. For these Fisher's inequality $b \geqslant v$ holds.*

(ii) *Semiregular group divisible designs for which $r-\lambda_1 > 0$, $rk-v\lambda_2 = 0$. For these designs $b \geqslant v-m+1$.*

(iii) *Singular group divisible designs for which $r-\lambda_1 = 0$. For these designs $b \geqslant m$.*

From (6.21) it follows that for a group divisible design

$$|NN'-I\theta| = (rk-\theta)(rk-v\lambda_2-\theta)^{m-1}(r-\lambda_1-\theta)^{v-m}. \qquad \ldots \quad (6.28)$$

Hence for a regular symmetrical GD design

$$(\det N)^2 = |NN'| = r^2(r^2-v\lambda_2)^{m-1}(r-\lambda_1)^{m(n-1)}. \quad \ldots \quad (6.29)$$

It follows that *for a regular symmetrical GD design* (i) *if m is even, then $r^2-v\lambda_2$ is a perfect square;* (ii) *if m is odd and n is even, then $r-\lambda_1$ must be a perfect square.*

The results given above were first obtained by Bose and Connor (1952), by a more direct method.

We have used the group divisible designs as illustration but it is clear that similar results can be proved for PBIB designs based on the other association schemes described in Section 4.

(f) We shall now examine for two class association schemes and PBIB designs, the consequences of the integral or non-integral nature of $\sqrt{\Delta}$, and the equality or inequality of the characteristic roots α_1 and α_2 of NN' (other than the simple root rk). These results are due to Connor and Clatworthy (1954).

The parameters n_1 and n_2 are both integers. If n_2 is zero any two treatments will be first associates, and the corresponding design will become a BIB. We shall exclude this trivial case and assume that n_1 and n_2 are non-zero positive integers. We shall set

$$\eta = (n_1-n_2)+\gamma(n_1+n_2). \quad \ldots \quad (6.30)$$

(i) *If Δ is not an integral square $\alpha_1 = \alpha_2$. Conversely if $\alpha_1 = \alpha_2$, Δ may or may not be an integral square.* From (2.2), (6.19), (6.20) and (6.30)

$$\alpha_1 = \frac{v-1}{2} - \frac{\eta}{2\sqrt{\Delta}}, \quad \alpha_2 = \frac{v-1}{2} + \frac{\eta}{2\sqrt{\Delta}}. \quad \ldots \quad (6.31)$$

If Δ is not an integral square, we nust have $\eta = 0$, since α_1 and α_2 must be rational. Hence $\alpha_1 = \alpha_2$.

Conversely if $\alpha_1 = \alpha_2$, then $\eta = 0$, but this does not put any restriction on Δ. Consider for example the cyclic association scheme with parameters (4.39), (4.40). In this case

$$\alpha_1 = \alpha_2 = 6, \qquad \Delta = 12.$$

Again for the L_2 association scheme with parameters (4.24), (4.25) taking $k = 3$, we have

$$\alpha_1 = \alpha_2 = 4, \qquad \Delta = 9.$$

(ii) *If $\alpha_1 = \alpha_2$ (which is necessarily the case when Δ is not an integral square)*

$$p^1_{12} = p^2_{12} = t, \quad \ldots \quad (6.32)$$

$$\alpha_1 = \alpha_2 = n_1 = n_2 = (v-1)/2 = 2t, \quad \ldots \quad (6.33)$$

$$v = \Delta = 4t+1. \quad \ldots \quad (6.34)$$

Since $\alpha_1 = \alpha_2$, we have $\eta = 0$ or

$$\gamma = p_{12}^2 - p_{12}^1 = \frac{n_2 - n_1}{n_2 + n_1} < 1,$$

whence $\gamma = 0$, $n_2 = n_1$, and $p_{12}^1 = p_{12}^2 = t$ (say). The other results easily follow.

(iii) *If v is even, Δ must be an integral square and η given by* (6.30), *must be a non-zero integer divisible by $\sqrt{\Delta}$.* This follows from (6.31) by noting that $\eta/\sqrt{\Delta}$ must be an odd integer.

(iv) *If v is odd, then either $\eta = 0$, $\alpha_1 = \alpha_2$* (in which case (6.32), (6.33), (6.34) hold) *or η is a non-zero even integer, and Δ is an integral square.* This follows by noting that from (6.31), $\eta/2\sqrt{\Delta}$ must be integral.

7. Partial geometries and the corresponding PBIB designs

A partial geometry (r, k, t) is a system of points and lines, and a relation of incidence between them satisfying the following axioms:

A1. Any two distinct points are incident with not more than one line.

A2. Each point is incident with r lines.

A3. Each line is incident with k points.

A4. If the point P is not incident with the line l, there are exactly t lines $(t \geqslant 1)$ which are incident with P, and also incident with some point incident with l.

Clearly $1 \leqslant t \leqslant k$, $1 \leqslant t \leqslant r$.

(a) If there were two distinct lines l and m each incident with two distinct points P_1 and P_2, then A1 would be contradicted. Hence

A′1. Any two distinct lines are incident with not more than one point.

Given a partial geometry (r, k, t), there exists a dual partial geometry (k, r, t), obtained by calling the points of the first, the lines of the second; and the lines of the second the points of the first.

The above follows by noting the duality of A1 and A′1, the duality of A2 and A3, and the self-dual nature of A4.

For convenience we may introduce the ordinary geometric language. Thus if a point is incident with a line we say that the point lies on the line, or is contained in the line, and the line passes through the point. If two points are incident on a line we speak of the line as joining the two points. If a point is incident with each of two lines, we say that the lines intersect in that point. With this language A4 may be rephrased as

A4. Through any point P not lying on a line l, there pass exactly t lines intersecting l.

It is easy to show that the number of points v, and the number of lines b in a partial geometry (r, k, t) are given by

$$v = k[(r-1)[k-1)+t]/t \qquad \dots \quad (7.1)$$

$$b = r[(r-1)(k-1)+t]/t. \qquad \dots \quad (7.2)$$

(b) Partial geometries are isomorphic with a class of PBIB designs. Points of the geometry may be called treatments and lines of the geometry may be called blocks. The relation of incidence now becomes the relation of a treatment being contained in a block. Two treatments can be called first associates if they occur together in a block and second associates if they do not occur in a block. Using the axioms A1 to A4 one can now prove that the association relations thus defined, satisfy the conditions (a), (b) of Section 1, and the condition (c′) of Section 3. The parameters of the two class association scheme thus obtained are

$$v = k[(r-1)(k-1)+t]/t, \; n_1 = r(k-1), \; n_2 = (r-1)(k-1)(k-t)/t, \qquad \dots \quad (7.3)$$

$$P_1 = \begin{pmatrix} (t-1)(r-1)+k-2 & (r-1)(k-t) \\ (r-1)(k-t) & (r-1)(k-t)(k-t-1)/t \end{pmatrix}, \qquad \dots \quad (7.4)$$

$$P_2 = \begin{pmatrix} rt & r(k-t-1) \\ r(k-t-1) & \dfrac{(r-1)(k-1)(k-t)}{t} - r(k-t-1)-1 \end{pmatrix}. \qquad \dots \quad (7.5)$$

This association scheme may be called the *geometric association scheme* with characteristics (r, k, t).

Thus a partial geometry is equivalent to a PBIB design with parameters of the second kind

$$v, b, r, k, \lambda_1 = 1, \; \lambda_2 = 0, \qquad \dots \quad (7.6)$$

the parameters of the association scheme being given by (7.3), (7.4), (7.5).

(c) Bose and Clatworthy (1955) considered two class PBIB designs with $r < k$, $\lambda_1 = 1$, $\lambda_2 = 0$. From the results of Section 6, it follows that for such designs the matrix Π^* given by (6.13) has a zero characteristic root, if we take $\lambda_1 = 1$, $\lambda_2 = 0$. Hence from (6.14)

$$rp_{12}^1-(r-1)p_{12}^2 = r(r-1). \qquad \dots \quad (7.7)$$

Using this relation, and the identities (1.3), (2.1), (2.2), (2.3), (2.4) they showed that the parameters of the design must be given by (7.3), (7.4), (7.5), (7.6). This raises the interesting question, whether the design is a partial geometry. The answer is in the affirmative. Since the axioms A1, A2, A3 are evidently satisfied, it only remains to show that A4 is also satisfied.

Let K be the set of k treatments contained in a particular block, and let $\bar{K}$ be the set of remaining $v-k$ treatments. Let $g(x)$ denote the number of treatments in $\bar{K}$ which have exactly x first associates in K. Then

$$\sum_{x=0}^{k} g(x) = v-k = k(k-1)(r-1)/t. \quad \ldots \quad (7.8)$$

By counting the number of pairs (P, Q), where P is a treatment in K, Q is a treatment in $\bar{K}$, and P and Q are first assciates, we see that

$$\sum_{x=0}^{k} x\, g(x) = k(n_1-k+1) = k(r-1)(k-1). \quad \ldots \quad (7.9)$$

Similarly by counting the number of triplets (P_1, P_2, Q) where P_1P_2 is an ordered pair of distinct treatments in K, and Q is a treatment in $\bar{K}$ which is a first associate of both K_1 and K_2, we get

$$\sum_{x=0}^{k} x(x-1)g(x) = k(k-1)(p^1_{11}-k+2) = k(k-1)(t-1)(r-1). \quad \ldots \quad (7.10)$$

A simple calculation shows that $\bar{x}$, the average value of x, is

$$\bar{x} = \Sigma x g(x)/\Sigma g(x) = t, \quad \ldots \quad (7.11)$$

and

$$\text{var } x = \sum_{x=0}^{k} g(x)(x-t)^2 = 0. \quad \ldots \quad (7.12)$$

Hence x must always have the value t. This is equivalent to the axiom A4. *Hence a PBIB design with r replications, block size k, $\lambda_1 = 1$, $\lambda_2 = 0$, is a partial geometry* (r, k, t) if $r < k$.

(d) One may ask whether a partial geometry (r, k, t) exists for all values of r, k, t. Now for the corresponding PBIB design the multiplicity α_1 of the characteristic root θ_1 of the incidence matrix NN' is given by (6.19). Substituting for n_1, n_2, v and Δ from (6.15), (7.3), (7.4) and (7.5), we have

$$\alpha_1 = \frac{rk(r-1)(k-1)}{t(k+r-t-1)}. \quad \ldots \quad (7.13)$$

Hence *a necessary condition for the existence of a partial geometry* (r, k, t) *is that the number α_1 given by* (7.13) *is a positive integer*. For example if $r = 3$, $t = 1$ then the only possible values of k are $k = 2, 3, 5$ and 11. The cases $k = 2, 3, 5$ are possible, but a rather lengthy combinatorial argument (Bose and Clatworthy, 1955) shows the case $k = 11$ to be impossible.

(e) We shall now consider some examples of partial geometries.

(i) Consider the lattice designs described in Section 4(d) based on the Latin square scheme L_r. The parameters of these designs are given by (4.27), (4.28), (4.29), (4.30). If we consider the treatments as points and blocks as lines, then axioms A1, A2, A3 are obviously satisfied. Now it follows from the method of derivation of the lattice designs from orthogonal Latin squares, that any two blocks of different replications have exactly one treatment in common. Suppose there is a treatment α and a block B^* of the i-th replication not containing α. Then there will be exactly one block in each replication which contains α. Let $B_1, B_2, \ldots, B_i, \ldots, B_r$ be these blocks. Then B_i has no treatment in common with B^*, whereas the other blocks $B_1, \ldots, B_{i-1}, B_{i+1}, \ldots, B_r$ have each one treatment in common with B^*. This shows that axiom A4 is satisfied with $t = r-1$. Hence *a lattice design with parameters* (4.27), (4.28), (4.29), (4.30) *is a partial geometry* $(r, k, r-1)$. *The geometric association scheme with characteristics* $(r, k, r-1)$ is identical with the L_r association scheme. It is easy to verify that by putting $t = r-1$ the formulae (7.3), (7.4), (7.5) reduce to (4.27), (4.28), (4.29).

(ii) A BIB design D with b treatments, v blocks, k replications, block size r, and $\lambda = 1$ is clearly a partial geometry (k, r, r). The dual design D^*, viz. the singly linked block design, based on the SLB scheme corresponding to D is therefore the partial geometry (r, k, r) dual to the partial geometry (k, r, r). Hence a *singly linked block design with parameter* (4.19), (4.20), (4.21), (4.22) *is a partial geometry* (r, k, r). *The geometric association scheme, with characteristics* (r, k, r) *is identical with the SLB association scheme, and in particular the geometric association scheme with characteristics* $(2, m-1, 2)$ *is identical with the triangular association scheme.* It is easy to verify that the formulae (7.3), (7.4), (7.5) reduce to (4.19), (4.20), (4.21) by putting $t = r$, and to (4.6), (4.7), by putting $t = r = 2$, $k = m-1$.

(iii) We shall conclude by giving a rather less obvious example of a partial geometry. Consider an elliptic non-degenerate quadric Q_5, in the finite projective space $PG(5, p^n)$. This quadric is ruled by straight lines, called generators, but contains no plane. As shown by Primrose (1951) and Ray Choudhuri (1959; 1962a) there are $(s^3+1)(s+1)$ points and $(s^3+1)(s^2+1)$ generators in Q_5. Each generator contains $s+1$ points, and through each point pass s^2+1 generators, where $s = p^n$. If P is a point on Q_5 not contained in a generator l, then the polar 4-space of P intersects l in a single point P^*, and PP^* is a generator of Q_5. It can be verified by using theorems proved by Ray Chaudhuri that PP^* is the only generator through P, which intersects l. Considering the points and generators of Q_5 as points and lines, we have a partial geometry $(s^2+1, s+1, 1)$. The corresponding PBIB design was obtained by Ray Chaudhuri (1959, 1962), the special case $s = 2$ having been obtained earlier by Bose and Clatworthy (1955).

In the same way one can show that the configuration of points and generators on a non-degenerate quadric Q_4 in $PG(4, p^n)$ is a partial geometry $(s+1, s+1, 1)$ where $s = p^n$. The corresponding design was first obtained by Clatworthy (1952; 1954).

8. Pseudo-geometric association schemes and uniqueness and embedding theorems

A two class association scheme, which has the parameters (7.3), (7.4), (7.5) and for which the inequalities

$$1 \leqslant t \leqslant r, \qquad 1 \leqslant t \leqslant k$$

are satisfied is defined to be a *pseudo-geometric* association scheme with characteristics (r, k, t). Thus a pseudo-geometric association scheme with characteristics (r, k, t) has the same parameters as the geometric association scheme with characteristics (r, k, t), viz., the association scheme of a partial geometry (r, k, t). However an association scheme may be pseudo-geometric without being the association scheme of a partial geometry (r, k, t).

In particular a two class association scheme with characteristics $(r, k, r-1)$ may be called a *pseudo-L_r* scheme. It has the same parameters as an L_r association scheme. Similarly a two class association scheme with characteristics (r, k, r) may be called a *pesudo-SLB* scheme, and in particular a two class association scheme with characteristics $(2, m-1, 2)$ may be called a *pseudo-triangular* scheme.

(a) A subset of treatments of an association scheme G, any two of which are first associates is defined to be a *clique* of G. When G is the association scheme of a partial geometry there will exist in G a set Σ of distinguished cliques $K_1, K_2, \ldots, K_b$, corresponding to the lines of the geometry satisfying the following axioms :

A*1. Any two treatments of G which are first associates are contained in one and only one clique of Σ.

A*2. Each treatment of G is contained in r cliques of Σ.

A*3. Each clique of Σ contains k treatments of G.

A*4. If α is a treatment of G not contained in a clique K_i of Σ, there are exactly t treatments in K_i which are first associates of α, $(i = 1, 2, \ldots, b)$.

Hence any association scheme G in which there exists a set Σ of cliques K_1, $K_2, \ldots, K_b$, satisfying axioms A*1-A*4 is a geometric association scheme with characteristics (r, k, t), and we can base on it a PBIB design for which the second kind of parameters are $b, r, k, \lambda_1 = 1, \lambda_2 = 0$, where b is given by (7.2).

One may consider two class association schemes in which there exists a set Σ of cliques $K_1, K_2, \ldots, K_b$ satisfying one or more but not all of the axioms A*1, A*2, A*3, A*4, and investigate under what additional conditions they will be geometric association schemes, with characteristics (r, k, t). Thus the result obtained at the end of Section 7(c) may be rephrased as : *If there exists in an association scheme G, a set Σ of cliques $K_1, K_2, \ldots, K_b$, satisfying the axioms* A*1, A*2, A*3, *and if $r < k$, then G is a geometric association scheme with characteristics (r, k, t).*

Again one can prove (Bose, 1963) the following theorem : *Let G be a pseudo-geometric association scheme with characteristics (r, k, t).* If it is possible to find in G a set Σ of cliques $K_1, K_2, \ldots, K_b$, satisfying axioms A*1 and A*2 and if $r < k$, then G is a *geometric association scheme.*

(b) Generalizing a result of Bruck (1963) regarding the *pseudo-L_r* scheme, Bose (1962; 1963) proved the following result :

Let G be a pseudo-geometric scheme with characteristics (r, k, t), *for which*

$$k > p(r, t) = \tfrac{1}{2}[r(r-1)+t(r+1)(r^2-2r+2)] \qquad \dots \quad (8.1)$$

then G is a geometric association scheme.

In other words if (8.1) is satisfied, then we can find a set of cliques in G satisfying the axioms A*1, A*2, A*3, A*4. Taking these cliques to be lines, and the treatments to be points we get a partial geometry (r, k, t).

(c) Consider the special case $r = t = 2$, $k = m-1$ of the theorem of subsection (b). Then $p(r, t) = 7$. Remembering the results of Section 7(e)(ii), it follows *that a pseudo-triangulrar association scheme,* i.e., *an association scheme with parameters* (4.6), (4.7), *must be a triangular association scheme if* $m > 8$. This result is due to Connor (1954), who expressed it by saying that the *triangular association scheme is unique* if $m > 8$.

Shrikhande (1959a) proved the uniqueness of the triangular association scheme for $m = 5, 6$; and Chang (1959) and Hoffman (1960) proved the same for $m = 7$.

Both Hoffman (1960) and Chang (1960) have shown that for $m = 8$, the parameters (4.6), (4.7) do not completely determine the association scheme. There are three other possible schemes with the same parameters besides the triangular. This may be expressed as follows :

There are four non-isomorphic two class association schemes (including the triangular) with parameters

$$v = 28,\ n_1 = 12,\ n_2 = 15,\ P_1 = \begin{pmatrix} 6 & 5 \\ 5 & 10 \end{pmatrix},\ P_2 = \begin{pmatrix} 4 & 8 \\ 8 & 6 \end{pmatrix}. \qquad \dots \quad (8.2)$$

Consider a BIB design with parameters

$$v^* = \tfrac{1}{2}(m-1)(m-2),\quad b^* = \tfrac{1}{2}m(m-1),\ r^* = m,\ k^* = m-2,\ \lambda^* = 2 \quad \dots \quad (8.3)$$

Hall and Connor (1953) have shown that *if this design exists then it can be embedded in a symmetric BIB design with parameters*

$$v_0 = b_0 = \tfrac{1}{2}m(m-1)+1,\quad r_0 = k_0 = n,\quad \lambda_0 = 2. \qquad \dots \quad (8.4)$$

One important consequence of Hall and Connor's embedding theorem is that the non-existence of (8.4) implies the non-existence of (8.3). For example if $m = 10$, it was shown independently by Schutzenberger (1949) and Shrikhande (1950a, 1950b) that the BIB design $v_0 = b_0 = 46$, $r_0 = k_0 = 10$, $\lambda_0 = 2$ does not exist. This implies the non-existence of the BIB design $v^* = 36$, $b^* = 45$, $r^* = 10$, $k^* = 10$, $\lambda^* = 2$.

Hall and Connor's proof does not cover the case $m = 8$, for which Connor (1951, 1952) separately showed the non-existence of (8.3). Shrikhande (1960), has given a very simple proof of the Hall and Connor theorem, when $m \neq 8$, using the

uniqueness of the triangular scheme. It is interesting to observe that $m = 8$, the case not covered in Hall and Connor's entirely different proof is exactly the case when the parameters (4.6), (4.7) do not uniquely characterize the scheme as triangular.

(d) Now consider the special case $t = r-1$, of the theorem of sub-section (b). Then

$$p(r, t) = \tfrac{1}{2}(r-1)(r^3-r^2+r+2). \qquad \ldots \quad (8.5)$$

In particular if $r = 2$, $p(r, t) = 4$. Remembering the results of Section 7(e)(i), it follows that a *pseudo-L_r scheme*, i.e., an association scheme with parameters (4.27), (4.28), (4.29) *is an L_r scheme provided that*

$$k > \tfrac{1}{2}(r-1)(r^3-r^2+r+2). \qquad \ldots \quad (8.6)$$

In other words if (8.6) is satisfied, and a two class association scheme with parameters (4.27), (4.28), (4.29) exists, then we can arrange the k^2 treatments in a $k \times k$ square scheme, and find a corresponding set of $r-2$ mutually orthogonal Latin squares, such that two treatments are first associates if they occur together in the same row or column of the square scheme, or correspond to the same symbol of one of the Latin squares, and are second associates otherwise. This result may be expressed by saying that *the L_r association scheme is unique when $r=2$ and unique up to type when $r > 2$, if* (8.6) *holds.* In the case $r > 2$, it is necessary to add the words unique up to type because in general there will exist many non-isomorphic sets of $r-2$ mutually orthogonal Latin squares. The uniqueness result for the L_2 scheme is due to Mesner (1956) and Shrikhande (1959b), and for the general case L_r (in a different language) to Bruck (1963). A slightly weaker result for the uniqueness of the L_r scheme was proved earlier by Mesner (1956), the bound (8.6) due to Bruck being sharper than Mesner's bound.

(e) Again let us consider the special case $t = r$ of the theorem of section (b). Then

$$p(r, t) = \tfrac{1}{2}r(r^3-r^2+r+1). \qquad \ldots \quad (8.7)$$

Remembering the results of Section 7(e)(ii), it follows that a *pseudo-SLB scheme*, i.e., an association scheme with parameters (4.19), (4.20), (4.21) *is an SLB scheme if*

$$k > \tfrac{1}{2}(r^3-r^2+r+1). \qquad \ldots \quad (8.8)$$

In other words if (8.8) is satisfied, and a two class association scheme exists with parameters (4.19), (4.20), (4.21), then there will exist a singly linked block design D^* (dual of some BIB design D with $\lambda = 1$), such that two treatments are first associates if they occur together in the same block of D^* and second associates if they occur together in different blocks. This result may be expressed by saying that *the SLB association scheme is unique up to type if* (8.8) *is satisfied.* It is necessary to add the words up to type because there exist in general non-isomorphic BIB designs D, with the same parameters, whose duals D^* are in consequence non-isomorphic.

(f) It is now clear that the result of sub-section (b) can be viewed as a generalized uniqueness theorem. We may say that *if there is a two class association scheme with parameters* (7.3), (7.4), (7.5), then if

$$k > \tfrac{1}{2}[r(r-1)+t(r+1)(r^2-2r+2)] \qquad \ldots \quad (8.9)$$

then the association scheme is unique up to type, in the sense that there will exist a partial geometry (r, k, t) *such that two treatments will be first associates if and only if the corresponding points are incident with the same line of the geometry.*

(g) It is well known that there cannot exist more than $k-1$ mutually orthogonal Latin squares of order k. A set of $k-1$ mutually orthogonal Latin squares of order k is called a complete set.

Given a set S of $r-2$ mutually orthogonal Latin squares of order k, we may define

$$d = (k-1)-(r-2) = k-r+1 \qquad \ldots \quad (8.10)$$

to be the deficiency of the set. If there exist d new Latin squares, such that when added to the original set S of $r-2$ mutually orthogonal squares, all the squares of the extended set are orthogonal, then it would be possible to extend the set S to a complete set.

Shrikhande (1961) proved that *if $k \neq 4$, then a set of mutually orthogonal Latin squares of order k and deficiency* 2 *can be extended to a complete set.* Bruck (1963) generalized this result and showed that *if*

$$k > \tfrac{1}{2}(d-1)(d^3-d^2+d+2) \qquad \ldots \quad (8.11)$$

then a set of mutually orthogonal Latin squares of order k and deficiency d can be extended to a complete set.

We shall here prove the following generalized embedding theorem and derive from it the results of Shrikhande and Bruck.

Given a two class association scheme G with the parameters

$$v = k[(d-1)(k-1)+t]/t,\ n_1 = (d-1)(k-1)/t,\ n_2 = d(k-1) \qquad \ldots \quad (8.12)$$

$$P_1 = \begin{pmatrix} \dfrac{(d-1)(k-1)(k-t)}{t} - d(k-t-1)-1 & d(k-t-1) \\ d(k-t-1) & dt \end{pmatrix} \qquad \ldots \quad (8.13)$$

$$P_2 = \begin{pmatrix} (d-1)(k-t)(k-t-1)/t & (d-1)(k-t) \\ (d-1)(k-t) & (t-1)(d-1)+(k-2) \end{pmatrix} \qquad \ldots \quad (8.14)$$

and a PBIB design based on this association scheme, with r replications, block size k, and $\lambda_1 > \lambda_2$, then we can extend the design by adding new blocks containing the same treatments, in such a way that the extended design is a BIB design, with $r_0 = r+d(\lambda_1-\lambda_2)$

HOFFMAN, A. J. (1960) : On the uniqueness of the triangular association scheme. *Ann. Math. Stat.*, **31**, 492-497.

——— (1960) : On the exceptional case in a characterization of the arcs of a complete graph. *IBM J.*, **4**, 487-496.

HALL, M. and CONNOR, W. S. (1953) : An embedding theorem for balanced incomplete block designs. *Can. J. Math.*, **6**, 35-41.

MESNER, D. M. (1956) : An investigation of certain combinatorial properties of partially balanced incomplete block designs and association schemes, with a detailed study of designs of the Latin square and related types. *Michigan State University, Doctoral Thesis.*

NAIR, K. R. (1943) : Certain inequality relations among the combinatorial parameters of balanced incomplete block designs. *Sankhyā*, **6**, 255-259.

NAIR, K. R. and RAO, C. R. (1942) : A note on partially balanced incomplete block designs. *Science and Culture*, **7**, 568-569.

PRIMROSE, E. J. F. (1951) : Quadratics in finite geometries. *Proc. Camb. Phil. Soc.*, **47**, 299-304.

RAY CHAUDHURI, D. K. (1959) : On the application of geometry of quadrics to the construction of partially balanced designs and error correcting codes. *University of North Carolina, Doctoral Thesis.*

——— (1962a) : Some results on quadrics in finite projective geometry. *Can. J. Math.*, **14**, 129-138.

——— (1962b) : Application of geometry of quadrics for constructing PBIB designs. *Ann. Math. Stat.*, **33**, 1175-1186.

SCHUTZENBERGER, M. P. (1949) : A non-existence theorem for an infinite family of symmetrical block designs. *Ann. Eugen.*, **14**, 286-287.

SHRIKHANDE, S. S. (1950a) : Construction of partially balanced designs and related problems. *University of North Carolina, Doctoral Thesis.*

——— (1950b) : The impossibility of certain symmetrical balanced incomplete block designs. *Ann. Math. Stat.*, **21**, 106-111.

——— (1952) : On the dual of certain balanced incomplete block designs. *Biometrics*, **8**, 66-72.

——— (1959a) : On a characterization of the triangular association scheme. *Ann. Math. Stat.*, **30**, 39-47.

——— (1959b) : The uniqueness of the L_2 association scheme. *Ann. Math. Stat.*, **30**, 781-798.

——— (1960) : Relations between certain incomplete block designs. *Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling*, 388-395. Stanford University Press.

——— (1961) : A note on mutually orthogonal Latin squares. *Sankhyā*, **23**, Series A, 115-116.

THOMSON, W. A. (1954) : On the ratio of variances in the mixed incomplete block model. *University of North Carolina, Doctroral thesis.*

——— (1958) : A note on PBIB design matrices. *Ann. Math. Stat.*, **29**, 919-922.

YATES, F. (1936a) : A new method of arranging variety trials involving a large number of varieties. *J. Agri. Sc.*, **26**, 424-455.

——— (1936b) : Incomplete randomised blocks. *Ann. Eugen.*, **7**, 121-140.

Paper received : March, 1963.

CLUSTER VALUES OF SEQUENCES OF ANALYTIC FUNCTIONS

By J. L. DOOB

University of Illinois

SUMMARY. Let f_n be a function meromorphic on the unit disc and let Γ_n be a subset of the disc perimeter. Let V_n be a suitably defined cluster set of f_n on Γ_n. In 1933 the author derived relations between the cluster values of f_n and V_n sequences under the restriction that Γ_n was an arc. In the present paper these results are extended, with necessarily weaker conclusions, to the case in which Γ_n is a Lebesgue measurable set.

1. Introduction

Let f_n be a function meromorphic on the unit disc $\{z : |z| < 1\}$ and let Γ_n be a subset of the unit disc perimeter, Lebesgue measurable with measure $|\Gamma_n| > 0$. Let R be the range of the f_n sequence, that is the set of those values taken on by infinitely many functions of the sequence. Let $B_1^0\{(f_n, \Gamma_n)\}$ be the boundary limit set along the Γ_n sequence, that is the set of those values α for which there is a sequence of integers $a_1 < a_2 < \cdots$, and a sequence $\{\alpha_{a_n}, n \geqslant 1\}$ with limit α for which α_{a_n} is a limiting value (cluster value) of f_{a_n} at some point of Γ_{a_n}. Relations were found by the author (Doob, 1933) between R, B_1^0 and certain limiting values of the f_n sequence (those obtained considering f_n only near Γ_n in a certain sense) under the hypothesis that Γ_n is an open arc for all n. In this paper analogous results will be obtained in the general case.

The results will be applied to find relations between the cluster set at a boundary point of a function meromorphic on the unit disc, the local range of the function near the point, and the boundary cluster set of the function along a measurable subset Γ of the disc perimeter metrically dense at the point. The relations found generalize previous work (Doob, 1963) in which Γ is an arc and work by other authors (see Noshiro, 1960 for detailed references) who assumed that Γ includes a neighbourhood of the perimeter point in question except perhaps for a set of capacity 0 or measure 0. The case in which only the perimeter point itself is excluded is classical.

All our definitions and results will be invariant under linear transformations of the unit disc onto itself, that is, if L_n is such a transformation and if the (f_n, Γ_n) sequence is replaced by the $(f_n(L_n), L_n(\Gamma_n))$ sequence, none of the cluster sets or range sets will be affected by the change.

2. Generalized harmonic measure

If D is an open plane set with boundary D' of strictly positive capacity and if Γ is a plane set, define $\mu(\cdot, \Gamma, D)$, the 'reduite' of 1 on Γ in Brelot's terminology (1956), as the superharmonic function on D which is equal except on a set zero capacity to the lower envelope of the class of superharmonic functions on D which satisfy the following conditions.

(a) $v \geqslant 0$;

(b) $v(z) \geqslant 1$ if z is in $\Gamma \bigcap D$;

(c) $v(z) \geqslant 1$ if z is in the set which is the intersection with D of some neighbourhood of $\Gamma \bigcap D'$.

The function $\mu(\cdot, \Gamma; D)$ is harmonic on the set of inner points of D-$D \bigcap \Gamma$. The function $\mu(z, \cdot; D)$ is monotone increasing and countably subadditive. If $\Gamma \subset D'$, $\mu(z, \Gamma; D)$ is the outer harmonic measure of Γ relative to D, at z. If the closure of Γ is a compact subset of D, $\mu(\cdot, \Gamma; D)$ is the outer capacitary potential of Γ. In probability language, if Γ is a Borel set, $\mu(z, \Gamma; D)$ is the probability of the set of those Brownian paths from z which either meet Γ before leaving D or meet D' for the first time in a point of Γ. (If D is a disc, $\Gamma \bigcap D'$ need only be Lebesgue measurable.)

The function $\mu(\cdot, \Gamma; D)$ has value 1 on $\Gamma \bigcap D$ except possibly on a set of zero capacity, and has limit 1 along almost every Brownian path from a point of D to D' which meets D' in a point of Γ. If D is simply connected, the last assertion is equivalent to the assertion that $\mu(\cdot, \Gamma; D)$ has limit 1 at almost every (harmonic measure on D') point of $\Gamma \bigcap D'$ on approach in the potential-theoretic fine topology.

If D is the unit disc, we omit it from the notation, writing simply $\mu(z, \Gamma)$.

Let Γ be a Lebesgue-measurable subset of the unit disc perimeter, of strictly positive measure, and let H be the set of those points z at which $\mu(z, \Gamma) > c$ for some constant $c < 1$. Then each open component of H is simply connected and has a subset of Γ of strictly positive measure on its boundary. Moreover $\mu(\cdot, \Gamma; H) > 0$. For almost every point z of Γ some open component of H includes the interior (sufficiently close to z) of any angle formed by rays from z into the unit disc. This follows from Fatou's theorem. If g is a function on an open component H_0 of H, bounded and regular, g has a non-tangential limit at almost every (Lebesgue measure) point z of Γ which is the vertex of an angle whose points sufficiently near z, except for z itself, lie in H_0. Moreover, the supremum of $|g|$ is at most the maximum of the supremum of its cluster values on the part of the boundary of H_0 in the disc and the essential supremum of the non-tangential boundary limit function on Γ. (Here 'essential supremum' means as usual the supremum up to value on a set of measure 0.) These facts can be derived using the standard methods applied in discussing the Lusin-Privalov and Plessner theorems. Moreover, the same statements except that 'non-tangential' is replaced by 'fine' are true because the Lusin-Privalov-Plessner results are valid with these changes in the usual discussion (Doob, 1961).

The following lemma will be fundamental. Let Γ_n be a Lebesgue-measurable subset of the unit disc perimeter, with $\lim_{n\to\infty} |\Gamma_n| = 2\pi$, so that $\lim_{n\to\infty} \mu(\cdot, \Gamma_n) = 1$ uniformly on every compact subset of the unit disc. Let $H_n(\varepsilon)$ be the open component of the set $\{z : \mu(z, \Gamma_n) > 1-\varepsilon\}$ containing 0. Since every compact subset of the unit disc lies in $H_n(\varepsilon)$ for n sufficiently large, depending on the compact subset, the following lemma makes sense.

Lemma 2.1 : *Under the preceding hypotheses,* $\lim_{n\to\infty} \mu(\cdot,\Gamma_n; H_n(\varepsilon)) = 1$ *uniformly on every compact subset of the unit disc.*

It is sufficient to prove the assertion for a sufficiently sparse subsequence. Hence we can and shall suppose that $\Sigma_n(2\pi - |\Gamma_n|) < \infty$. Define $\Gamma'_n = \bigcap_{k\geqslant n} \Gamma_k$. Then $\Gamma'_1 \subset \Gamma'_2 \subset \dots$, $|\Gamma'_n| \to 2\pi$ and $\lim_{n\to\infty} \mu(\cdot, \Gamma'_n) = 1$ uniformly on every compact subset of the unit disc. It will be convenient to prove the lemma by probability methods. Consider those Brownian paths from 0, a path set of probability $\mu(0, \Gamma'_n)$, which meet the unit disc perimeter for the first time in a point of Γ'_n. For each value of n, $\mu(\cdot, \Gamma'_n) \to 1$ on almost every Brownian path from 0 to Γ'_n and we restrict our attention to those paths which meet the perimeter for the first time at a point of $\bigcup_n \Gamma'_n$ and for which the preceding limit relation is true simultaneously for all sufficiently large n. The set of these paths has probability 1. If $n \geqslant k$ the function $\mu(\cdot, \Gamma'_n)$ considered as a function of the Brownian path parameter on the closed interval from 0 to the time the path first meets Γ'_k, defining the function as 1 at the parameter interval endpoint, is continuous for sufficiently large n and increases with n to 1. Hence the convergence to 1 is uniform, and we conclude that the whole path to Γ'_k lies in $H_n(\varepsilon)$ (except for the endpoint on Γ'_k) for sufficiently large n. Thus we have proved that almost every Brownian path from 0 to the disc perimeter lies in $H_n(\varepsilon)$ (except for the endpoint on the perimeter) for sufficiently large n. Now $\mu(0, \Gamma_n; H_n(\varepsilon))$ is the probability measure of the set Λ_n of Brownian paths from the disc center which meet the boundary of $H_n(\varepsilon)$ for the first time in a point of Γ_n, and what we have just proved is that $\liminf_{n\to\infty} \Lambda_n$ includes almost all Brownian paths from the disc center. We infer that $\lim_{n\to\infty} \mu(\cdot, \Gamma_n; H_n(\varepsilon)) = 1$ at 0 and hence uniformly in every compact subset of the unit disc (Harnack inequality), as was to be proved.

3. Boundary limit sets of an (f_n, Γ_n) sequence

In the following, the function f_n is meromorphic on the unit disc and Γ_n is a subset of the unit disc perimeter of strictly positive Lebesgue measure. In Section 1 we defined a cluster set $B_1^0\{(f_n, \Gamma_n)\}$ of the f_n sequence along the Γ_n sequence. Several other such sets will now be defined. Let $B_2^0\{(f_n, \Gamma_n)\}$ be the set of those values α for which there are a sequence of integers $a_1 < a_2 < \dots$ and a sequence $\{\alpha_{a_n}, n \geqslant 1\}$ with limit α for which α_{a_n} is the limit of f_{a_n} along a nontangential sequence to some point of Γ_{a_n}. At each point of Γ_n choose an angle with vertex at the point, opening into the disc. If a point lies in several boundary sets Γ_n, the angle chosen may depend on n.

Define $B_3^0\{(f_n, \Gamma_n)\}$ like the cluster set just defined except that α_{a_n} is to be the limit o f_{a_n} along a sequence to a point of Γ_{a_n} where now the sequence is to lie in the specified (closed) angle for Γ_{a_n} at that point. Define $B_4^0\{(f_n, \Gamma_n)\}$ and $B^0\{(f_n, \Gamma_n)\}$ similarly except that for B_4^0 α_{a_n} is to be a fine cluster value, and for B^0 α_{a_n} is to be the fine limit of f_{a_n} at some point of Γ_{a_n} at which this function has a fine limit. (See Doob (1961) for a discussion of fine cluster values.) The set B^0 may be empty. The five cluster sets we have defined are all closed. Finally, define $B_1\{(f_n, \Gamma_n)\} = \bigcap B_1^0\{(f_n, \Gamma_n')\}$ where the intersection is over all Γ_n' sequences for which $\Gamma_n' \subset \Gamma_n$ and $|\Gamma_n - \Gamma_n'| = 0$. The other sets denoted without superscripts are defined similarly, and we omit (f_n, Γ_n) from the notation if there is no ambiguity. Since all the cluster sets defined are closed, there is a Γ_n' sequence making $B_1^0\{(f_n, \Gamma_n')\} = B_1\{(f_n, \Gamma_n)\}$, and the same is true for the other four types of cluster sets. Under our definitions, $B_1 \supset B_2 \supset B_3 \supset B_4 \supset B$. The first inclusion is obvious. The second is true because if f_n has a nontangential limit at almost every point of Γ_n, when n is sufficiently large, then $B_2 = B_3$ trivially, whereas if the f_n sequence does not have this property $B_2 = B_3 =$ (extended plane) according to Plessner's theorem. The inclusion $B_3 \supset B_4$ follows from the fact that the fine cluster set at z of any function from the unit disc to a compact metric space is not only, for almost every point z of the perimeter, included in the nontangential cluster set at z, but even in the nontangential cluster set obtained using a specified angle at each point z. (The weaker statement is Theorem 4.3 of Doob (1961); the proof of the theorem actually proves the stronger statement.) The inclusion $B_3 \supset B_4$ may be strict. In fact Constantinescu and Cornea (1960) have given an example of a meromorphic function on the unit disc which does not have a nontangential limit at any point of the perimeter but which has (in our terminology) a fine limit at almost every point of the perimeter. If f_n is this function for all n, and if Γ_n is arbitrary, B_3 is the extended plane, whereas B_4 can be made properly smaller if Γ_n is chosen suitably.

The theorems of the present paper are all assertions about the nature of the f_n sequence as related to the complement of a boundary limit set of the (f_n, Γ_n) sequence. The theorems are true for all ten of the boundary limit sets defined above, but are strongest with the choice B, for which the proof will be given. The relation $B_4 \supset B$ may be strict. According to the fine cluster value generalization of Plessner's theorem (Doob, 1961) either f_n has a fine limit at almost every point of Γ_n when n is sufficiently large, in which case $B_4 = B$ trivially, or not, in which case B_4 is the extended plane and B may be empty.

4. Relative cluster and range sets of an (f_n, Γ_n) sequence

If $0 < \omega \leqslant 1$, α will be called an ω-cluster value of an (f_n, Γ_n) sequence if there is a strictly increasing sequence $(a_n, n \geqslant 1)$ of positive integers and a sequence $(z_{a_n}, n \geqslant 1)$ in the unit disc for which

$$\lim_{n \to \infty} f_{a_n}(z_{a_n}) = \alpha, \quad \inf_n \mu(z_{a_n}, \Gamma_{a_n}) > 1-\omega. \qquad \ldots \quad (4.1)$$

The set of ω-cluster values will be denoted by $C(\{(f_n, \Gamma_n)\}, \omega)$ or simply by $C(\omega)$ if there is no ambiguity. We also define $C(0) = \bigcap_{\omega > 0} C(\omega)$. A point α is in $C(0)$ if and only if a_n, z_{a_n} can be found as above except that the second condition in (4.1) is replaced

by $\lim_{n\to\infty} \mu(z_{a_n}, \Gamma_{a_n}) = 1$. The set $C(0)$ is obviously closed. As ω increases $C(\omega)$ increases (wide sense). If (4.1) is modified by allowing equality in the second condition, the set $\overline{C}(\omega)$ so defined is closed and $C(\omega) \subset \overline{C}(\omega) \subset C(\omega')$ for $\omega' > \omega$. Thus $C(\omega)$ is an F_σ set. We shall prove that the part of this set not in B is open, under certain restrictions on the Γ_n sequence.

If $0 < \omega \leqslant 1$, α will be said to be in the ω-range of the (f_n, Γ_n) sequence if there is a strictly increasing sequence $\{a_n, n \geqslant 1\}$ of positive integers and a sequence $\{z_{a_n}, n \geqslant 1\}$ in the unit disc for which

$$f_{a_n}(z_{a_n}) = \alpha, \quad \inf_n \mu(z_{a_n}, \Gamma_{a_n}) > 1-\omega. \qquad \ldots \quad (4.2)$$

The ω-range will be denoted by $R(\{(f_n, \Gamma_n)\}, \omega)$ or by $R(\omega)$ if there is no ambiguity, and we also define $R(0) = \bigcap_{\omega>0} R(\omega)$. As ω increases, $R(\omega)$ increases (wide sense), and $R(\omega) \subset C(\omega)$.

If L_n is a linear transformation of the unit disc onto itself, the $(f_n(L_n), L_n(\Gamma_n))$ sequence has the same ω-cluster set and ω-range as the original sequence.

5. Cohesiveness

Let $\{(\Gamma_n, n \geqslant 1)\}$ be a sequence of Lebesgue-measurable subsets of the unit disc perimeter. Let $\{a_n, n \geqslant 1\}$ be a strictly increasing sequence of integers and let $\{z_{a_n}, n \geqslant 1\}$ be a sequence of points in the unit disc. Let $A_k(d)$ be the subset of the unit disc at hyperbolic distance less than d from z_k. If $0 < \omega \leqslant 1$ and if the conditions

$$\inf_n \mu(z_{a_n}, \Gamma_{a_n}) > 1-\omega, \ \lim_{d\to\infty} \inf_n \sup_{z \varepsilon A_{a_n}(d)} \mu(z, \Gamma_{a_n}) = 1 \qquad \ldots \quad (5.1)$$

are satisfied, the Γ_n sequence will be said to be weakly ω-cohesive for the z_{a_n} sequence. If $0 < \varepsilon < \mu(z_k, \Gamma_k)$, let $H_k(\varepsilon)$ be the set of values of z for which $\mu(z, \Gamma_k) > \varepsilon$ and let $A_k(d, \varepsilon)$ be the open component containing z_k of $A_k(d) \cap H_k(\varepsilon)$. We shall call the Γ_n sequence ω-cohesive for the z_{a_n} sequence if

$$\inf_n \mu(z_{a_n}, \Gamma_{a_n}) > 1-\omega, \ \lim_{d\to\infty} \inf_n \sup_{z \varepsilon A_{a_n}(d, 1-\omega')} \mu(z, \Gamma_{a_n}) = 1 \qquad \ldots \quad (5.2)$$

when $\omega-\omega'$ is strictly positive and sufficiently small.

We shall say that an (f_n, Γ_n) sequence is [weakly] ω-cohesive for the cluster value α if there is a z_{a_n} sequence for which the Γ_n sequence is [weakly] ω-cohesive and for which $f_{a_n}(z_{a_n}) \to \alpha$. If in addition to [weak] ω-cohesiveness for α with a specified z_{a_n} sequence there is a set G with the property that to each point β in G there corresponds a subsequence $\{b_n, n \geqslant 1\}$ of $\{a_n, n \geqslant 1\}$, a sequence $\{z_{b_n}, n \geqslant 1\}$ in the unit disc, and a positive number δ for which $[z_{b_n} \epsilon A_{b_n}(\delta)]$ $z_{b_n} \epsilon A_{b_n}(\delta, 1-\omega')$ if $\omega-\omega'$ is strictly positive and sufficiently small (uniformly as n varies) and $f_{b_n}(z_{b_n}) \to \beta$, we shall say that the (f_n, Γ_n) sequence is [weakly] ω-cohesive simultaneously for all points of G. The sequence will then be [weakly] ω-cohesive for all points of G starting from any point of G, not just from α.

If L_n is a linear transformation of the unit disc onto itself, the Γ_n sequence is [weakly] ω-cohesive for the z_{a_n} sequence if and only if the $L_n(\Gamma_n)$ sequence is for the $L_n(z_{a_n})$ sequence. Thus it will usually be possible to assume that $z_{a_n} = 0$ for all n. In this case the ω-cohesiveness condition becomes

$$\inf_n |\Gamma_{a_n}| > 1-\omega, \lim_{r\to 1} \inf_n \sup{}' \mu(z, \Gamma_{a_n}) = 1 \qquad \dots \quad (5.3)$$

where the primed supremum is for z in the open component containing 0 of $\{z : |z| < r\} \cap H_{a_n}(1-\omega')$ and the relation is to be true for $\omega-\omega'$ strictly positive and sufficiently small.

It will be useful to reformulate (5.3) in a different form. The sequence $\{\mu(\cdot, \Gamma_{a_n}), n \geqslant 1\}$ of harmonic functions is compact in the sense that every subsequence contains a further subsequence which is locally uniformly convergent. Let u be any limit function of a subsequence and define $H(\varepsilon)$ as the open component of $\{z : u(z) > \varepsilon\}$ containing 0. Then the second condition of (5.3) is equivalent to the condition that u have supremum 1 on $H(1-\omega)$ (for every limit function u). If $H(1-\omega)$ here is interpreted as the unit disc itself, the condition just formulated becomes that for weak ω-cohesiveness.

Changing each set Γ_n by a set of measure 0 has no effect on weak ω-cohesiveness or ω-cohesiveness. If $\Gamma_1 = \Gamma_2 = \cdots$ the Γ_n sequence is ω-cohesive for every z_{a_n} sequence with $\sup\limits_n |z_{a_n}| < 1$. If the Γ_n sequence is a sequence of arcs, it is ω-cohesive for every z_{a_n} sequence satisfying the first condition in (5.2).

6. A relation between B and $C(0)$

Theorem 6.1 : *For any (f_n, Γ_n) sequence, B contains the boundary of $C(0)$.*

This theorem was proved by Doob (1933) with B_1^0 instead of B, and the Γ_n sequence a sequence of arcs, and the following proof applies the same method. Let α be a point of $C(0)$ not in B. We shall prove that then some neighbourhood of α also lies in $C(0)$. Going to a subsequence and making linear transformations of the disc onto itself if necessary, we can suppose that $f_n(0) \to \alpha$, $|\Gamma_n| \to 2\pi$. If the f_n sequence is not normal, the sequence is not normal at some point of the unit disc. Then every value except possibly two in the extended plane is taken on by infinitely many members of the f_n sequence in an arbitrary neighbourhood of the point of non-normality. It follows that $C(0)$ is the extended plane and the assertion to be proved is verified in this case. If the f_n sequence is normal and has a nonconstant limit function (necessarily with value α at 0) some neighbourhood of α is in $C(0)$. Finally, suppose that $f_n \to \alpha$ locally uniformly. We can suppose that $\alpha \neq \infty$, replacing f_n by $1/f_n$ otherwise. Let $3d$ be the distance from α to B. If β is a point not in $C(0)$, with $|\alpha-\beta| < d$, β is at distance $> 2d$ from B. Let $H_n(\varepsilon)$ be the subset of the unit disc on which $\mu(\cdot, \Gamma_n) > 1-\varepsilon$. Then $H_n(\varepsilon)$ contains the origin if $\varepsilon > 0$ and n is large. If for every $\varepsilon > 0$ there is a sequence $\{(a_n, z'_{a_n}), n \geqslant 1\}$ for which $a_1 < a_2 < \dots$, $z'_{a_n} \epsilon H_{a_n}(\varepsilon)$, and $f_{a_n}(z'_{a_n}) \to \beta$, then $\beta \epsilon C(0)$, contrary to hypothesis. Hence there must be an ε and a K for which, for sufficiently large n,

$$|f_n - \beta|^{-1} \leqslant K, \quad \text{on} \quad H_n(\varepsilon).$$

But then, if n is sufficiently large,

$$|f_n-\beta|^{-1} \leqslant \mu(\cdot, \Gamma_n; H_n(\varepsilon))/2d+K[1-\mu(\cdot, \Gamma_n; H_n(\varepsilon))] \text{ on } H_n(\varepsilon) \quad \ldots \quad (6.1)$$

because the left side of (6.1) is a bounded subharmonic function on $H_n(\varepsilon)$ with continuous boundary values dominated by those of the harmonic function on the right at boundary points of $H_n(\varepsilon)$ in the unit disc, and with fine limits (as well as nontangential limits) existing and dominated by those of the harmonic function on the right at almost all points of Γ_n (see Section 2). When $n \to \infty$ the bracketed expression in (6.1) goes to 0, according to Lemma 2.1, at each point of the disc. Hence, setting the argument equal to 0 in (6.1) this inequality yields, when $n \to \infty$, $|\alpha-\beta|^{-1} \leqslant 1/2d$ which contradicts the conditions imposed on β. There can therefore be no such point β; some neighbourhood of α lies in $C(0)$, as was to be proved.

7. Relations between B and $C(\omega)$

In the following, the complement of a set X will be denoted by $\tilde{X}$.

Theorem 7.1 : *If the (f_n, Γ_n) sequence is ω-cohesive for the cluster value α in $\tilde{B}$, then $C(\omega)$ contains a neighbourhood of α and, if $\alpha \epsilon \tilde{C}(0)$, the (f_n, Γ_n) sequence is even simultaneously ω-cohesive for a neighbourhood of α.*

According to this theorem, the set of ω-cluster values in $\tilde{B}$ for which the (f_n, Γ_n) sequence is ω-cohesive is open. The simplest application of the theorem is to the case when the (f_n, Γ_n) sequence is ω-cohesive for every value in $C(\omega)$. This is true for example, irrespective of the f_n sequence, when Γ_n is an arc for all n, the case treated by Doob (1933). In this case $C(\omega) \cap \tilde{B}$ is open.

Theorem 7.1 is a trivial consequence of Theorem 6.1 if $\alpha \epsilon C(0)$. Hence we shall exclude this case in the following proof. Replacing the (f_n, Γ_n) sequence by a subsequence and making linear transformations of the unit disc onto itself, if necessary, we can assume that the Γ_n sequence is ω-cohesive for the sequence 0, 0, ... and that

$$\inf_n |\Gamma_n| > 1-\omega, \quad \lim_{n\to\infty} f_n(0) = \alpha, \quad \lim_{n\to\infty} \mu(\cdot, \Gamma_n) = u, \quad \sup_H u = 1. \quad \ldots \quad (7.1)$$

Here u is harmonic in the unit disc, the convergence to u is locally uniform, and $H = \{z : u(z) > 1-\omega\}$. (a) If the restriction of the f_n sequence to H is normal, and if some limit function of the sequence on H is not identically α, the restriction of the sequence to the part of H in any neighbourhood of 0 has some neighbourhood of α as limit set. Hence the (f_n, Γ_n) sequence is ω-cohesive simultaneously for the points of a neighbourhood of α. (b) If the restriction of the f_n sequence to H converges locally uniformly to α, $\alpha \epsilon C(0)$ because of the last relation in (7.1). We have already treated and excluded this case. (c) Finally, if the restriction of the f_n sequence to H is not normal, there is a point of non-normality in H and we deduce at once that $C(\omega)$ is the extended plane, and in fact that the (f_n, Γ_n) sequence is simultaneously ω-cohesive for all the points of the extended plane.

Theorem 7.2 : *If the (f_n, Γ_n) sequence is weakly ω-cohesive for the cluster value α in $\tilde{B}$ and if $C(\omega)$ does not contain a neighbourhood of α, then $C(\omega')$ is the extended plane for some $\omega' < 1$, and in fact the (f_n, Γ_n) sequence is simultaneously weakly ω'-cohesive for all the points of the extended plane.*

The proof follows that of Theorem 7.1. Again we can exclude the case $\alpha \epsilon C(0)$ because the theorem becomes trivial in that case. We follow through steps (a) and (b) of the preceding proof, interpreting H as the unit disc. These steps need no change. In step (c) we can only conclude under the present hypotheses that, if z' is a point of non-normality of the f_n sequence on the unit disc and if ω' is so large that $\inf_n \mu(z', \Gamma_n) > 1-\omega'$, the (f_n, Γ_n) sequence is simultaneously weakly ω-cohesive for every point of the extended plane.

We observe that in both this and the previous theorem there is a uniformity of the simultaneous cohesiveness in the sense that (see the definition in Section 5) the number d can be chosen the same for all points of the set G in question.

8. The range of a meromorphic function

Theorem 8.1 : *Let f be a meromorphic function on the unit disc and let Γ be a Lebesgue-measurable perimeter set. Let D be an open set of the extended plane with the property that for some subset Γ' of Γ with $|\Gamma-\Gamma'| = 0$, no point of D is the fine limit of f at a point of Γ'. Let G be the subset of D not in the range of f. Then*

$$\mu(f(z), G; D) \leqslant 1-\mu(z, \Gamma) \quad if \quad f(z) \epsilon D-G. \qquad \ldots \quad (8.1)$$

This theorem is well known in various forms, and we shall therefore only sketch a proof. We observe that if G has zero capacity (8.1) becomes trivial because the left side vanishes, whereas if G has strictly positive capacity f takes the unit disc into a hyperbolic Riemann surface and so has a fine limit at almost every point of the disc perimeter (see Doob (1961) or, in a different terminology, Constantinescu and Cornea (1960)). Thus in the only interesting case f has a fine limit at almost every point of Γ.

We can assume in proving (8.1) that z is chosen with $f'(z) \neq 0$. Suppose that D is as stated. (For the following type of reasoning see Doob (1961)). Then almost all those Brownian paths from $f(z)$ leaving D first in a point of G must have as inverse images under f Brownian paths from z to points of the disc perimeter in $\tilde{\Gamma}$, except that the latter paths may correspond only to initial segments of the former ones. This fact implies the truth of (8.1), in view of the remarks in Section 2 on Brownian paths.

Theorem 8.2 : *For every strictly positive $\gamma < 2\pi$ and strictly positive $r < 1$ there is an absolute constant $\delta(\gamma, r)$, where $\lim_{\gamma \to 2\pi} \delta(\gamma, r) = 0$, with the property that if f is meromorphic on the unit disc, if $f(0) = 0$, and if at almost every point of a Lebesgue measurable subset Γ of the disc perimeter where f has a fine limit the limit has modulus $\geqslant 1$, then the capacity relative to the unit disc of the subset G_r of $\{z : |z| < r\}$ not in the range of f is $\leqslant \delta(|\Gamma|, r)$.*

Taking D in Theorem 8.1 as the unit disc and applying that theorem we find that

$$\mu(0, G_r; D) \leqslant \mu(0, G; D) \leqslant 1-|\Gamma|. \qquad \ldots \quad (8.2)$$

This inequality implies the truth of the theorem, because the first term in (8.2) is the value at 0 of the capacitary potential of G_r relative to the unit disc.

9. Relations between B, $C(0)$, $R(0)$

We consider again a function boundary-set sequence $\{f_n, \Gamma_n\}$ and its associated cluster and range sets.

Theorem 9.1 : *For any (f_n, Γ_n) sequence, the set $C(0)\cap\tilde{R}(0)\cap\tilde{B}$ has capacity* 0.

We first prove the theorem with $R(0)$ replaced by the in general larger set R. Let D_0 be an open component of $\tilde{B}$. According to Theorem 6.1, if D_0 contains a point of $C(0)$ then $D_0 \subset C(0)$. We show that in the latter case R includes D_0 except perhaps for a set of zero capacity. Let α be a finite point of D_0. There are then a function subsequence $\{f_{a_n}, n \geqslant 1\}$ and a point sequence $\{z_{a_n}, n \geqslant 1\}$ in the unit disc for which

$$\lim_{n\to\infty} f_{a_n}(z_{a_n}) = \alpha, \quad \sum_n [1-\mu(z_{a_n}, \Gamma_{a_n})] < \infty. \qquad \ldots \quad (9.1)$$

We can suppose that $f_{a_n}(z_{a_n}) \epsilon D_0$ for all n. Let B_n be the set of cluster values of f_n at the points of Γ_n, and let D be an open set containing α whose closure is a subset of D_0. It will be sufficient to prove that R includes all of D except perhaps a set of zero capacity. The set $B_{a_n}\cap D$ is empty if n is sufficiently large. If $G_{k\delta}$ is the set of points of D not in the range of f_k and at distance $> \delta$ from α,

$$\mu(f_{a_n}(z_{a_n}), G_{a_n\delta}; D) \leqslant 1-\mu(z_{a_n}, \Gamma_{a_n}),$$

according to Theorem 8.1. The function $\mu(\cdot, G_{a_n\delta}; D)$ is harmonic on the disc of center α, radius, δ, and according to a Harnack theorem there is a constant c for which

$$\mu(\alpha, G_{a_n\delta}; D) \leqslant c\mu(f(z_{a_n}), G_{a_n\delta}; D) \qquad \ldots \quad (9.2)$$

if n is so large that $|f(z_{a_n})-\alpha| < \delta/2$. Then for large j

$$\mu(\alpha, \bigcup_j^{\infty} G_{a_n\delta}; D) < c \sum_j^{\infty} [1-\mu(z_{a_n}, \Gamma_{a_n})] \qquad \ldots \quad (9.3)$$

and the sum is near 0 for large j. Hence $\mu(\alpha, \limsup_{n\to\infty} G_{a_n\delta}; D) = 0$, so that the indicated limit superior has zero capacity. Since the result is true for all $\delta > 0$, $\limsup_{n\to\infty} G_{a_n 0}$ has zero capacity, that is, R includes all of D except possibly a set of zero capacity. We have now proved the assertion of the theorem with $R(0)$ replaced by R. To prove the theorem as stated, suppose that (9.1) is true, let ε be a strictly positive number and let $H_k(\varepsilon)$ be the open component of the set $\{z : \mu(z, \Gamma_k) > 1-\varepsilon\}$ containing z_k. We suppose again that n is so large that $H_{a_n}(\varepsilon)$ exists. Applying linear transformations of the unit disc onto itself if necessary we can suppose that $z_{a_n} = 0$, $n \geqslant 1$.

According to Lemma 2.1, if $0 < \varepsilon < 1$,

$$\lim_{n\to\infty} \mu(z, \Gamma_{a_n}; H_{a_n}(\varepsilon)) = 1,$$

uniformly on every compact subset of the unit disc. The set $H_{a_n}(\varepsilon)$ is simply connected. Mapping it one-to-one and conformally onto the unit disc and applying the part of the theorem already proved to the transformed function-set sequence we find that, for $\varepsilon_0 > \varepsilon$, $C(0) \cap \tilde{R}(\varepsilon_0) \cap \tilde{B}$ has zero capacity. Since ε is an arbitrary strictly positive number, Theorem 9.1 follows.

10. Relations between B, $C(\omega)$, $R(\omega)$

Theorem 10.1 : *If the (f_n, Γ_n) sequence is ω-cohesive for the cluster value α in $\tilde{B}$, then $R(\omega)$ contains a neighbourhood of α except possibly for a set of zero capacity.*

Replacing the (f_n, Γ_n) sequence by a subsequence and making linear transformations of the unit disc onto itself, if necessary, we can assume that the Γ_n sequence is ω-cohesive for the sequence $0, 0, \ldots$ and that (7.1) is true. Let D be the open component of $\tilde{B}$ containing α. If $\alpha \epsilon C(0)$ then $D \subset C(0)$ and $D \cap \tilde{R}(0)$ has zero capacity according to Theorems 6.1 and 9.1. Thus the theorem is true in this case. If we exclude this case and use the fact that $\sup_H u(z) = 1$, we find that no subsequence of the f_n sequence can converge locally uniformly to α on H. We conclude that if restriction to H of the f_n sequence is normal $R(\omega)$ includes a neighbourhood of α because no limit function can be identically α. If the restriction to H of the f_n sequence is not normal, there is a point of non-normality in H and we conclude that $R(\omega)$ is the extended plane less at most two points.

The proof of Theorem 10.1 is now complete. Actually we have proved a stronger result. In the first place we have proved that the f_n sequence has range a neighbourhood of α close to any sequence for which the (f_n, Γ_n) sequence is ω-cohesive for α (the precise statement is the analogue for ranges of simultaneous ω-cohesiveness). In the second place we have proved that if D is any open component of $\tilde{B}$ then either $D \subset C(0)$ and then $\tilde{R}(0) \cap D$ has zero capacity, or $D \cap C(0)$ is empty and in that case if $0 < \omega \leqslant 1$ the set of values of α in D for which the (f_n, Γ_n) sequence is ω-cohesive is open and is in $R(\omega)$ except possibly for two points. If there is even one exceptional point, $R(\omega)$ is the extended plane less at most two points.

Theorem 10.2 : *Suppose that the (f_n, Γ_n) sequence is weakly ω-cohesive for the cluster value α in $\tilde{B}$. If $R(\omega)$ does not contain a neighbourhood of α there is an $\omega' < 1$ such that $R(\omega')$ includes the open component of $\tilde{B}$ containing α except possibly for a set of zero capacity.*

Applying a now-familiar argument, we can assume that (7.1) is true, where H is now interpreted to be the unit disc. The theorem is trivial if $\alpha \epsilon C(0)$ because of Theorem 9.1, so we exclude this case below. Then no subsequence of the f_n sequence can converge locally uniformly to α. If the f_n sequence is normal no limit function is identically α, so $R(\omega)$ contains a neighbourhood of α. If the f_n sequence is not normal, $R(\omega')$ is the extended plane less at most two points, for some $\omega' < 1$, depending on the location of the points of non-normality.

11. Application to local properties of a meromorphic function

Let f be a function meromorphic on the unit disc, let Γ be a Lebesgue-measurable perimeter subset and suppose that every neighbourhood of 1 contains a subset of Γ of strictly positive measure. Let $B_1^0(f, \Gamma)$ be the boundary limit set of f at the point 1 along Γ, that is, the set of those values α for which there is a sequence $\{\alpha_n, n \geqslant 1\}$ with limit α such that α_n is a cluster value of f at some point z_n of Γ and $\lim_{n\to\infty} z_n = 1$. Define $B_2^0(f, \Gamma)$ in the same way except that α_n is a nontangential cluster value, $B_4^0(f, \Gamma)$ in the same way except that α_n is a fine cluster value, $B^0(f, \Gamma)$ in the same way except that α_n is the fine limit of f at some point z_n of Γ where this fine limit exists. If for each point of Γ an angle is chosen, opening into the unit disc, with vertex at that point, and if α_n above is a cluster value of f at z_n along a sequence lying in the angle assigned to z_n, the boundary cluster set so defined will be denoted by $B_3^0(f, \Gamma)$. The dependence on the specified angles is omitted from the notation. Finally, $B_1(f, \Gamma)$ is defined as $\bigcap B_1^0(f, \Gamma')$ where the intersection is over all Lebesgue-measurable subsets Γ' of Γ with $|\Gamma - \Gamma'| = 0$. The sets $B_2(f, \Gamma)$ and so on are defined analogously. These ten sets are closed and each set denoted with a superscript reduces to the corresponding one denoted without a superscript, for a proper choice of Γ'. Finally, dropping (f, Γ) from the notation, as we usually shall do when there is no ambiguity, $B_1 \supset B_2 \supset B_3 \supset B_4 \supset B$. The results to be discussed here involve the complement of a boundary cluster set of f at 1. The results are correct for any of the ten choices, but are strongest when the smallest set B is chosen and will be proved with this choice (see the corresponding discussion in Section 3).

If $0 < \omega < 1$, a value α will be called an ω-cluster value of the pair (f, Γ) at 1 if there is a sequence $\{z_n, n \geqslant 1\}$ of points in the unit disc for which

$$\lim_{n\to\infty} z_n = 1, \qquad \lim_{n\to\infty} f(z_n) = \alpha, \qquad \inf_n \mu(z_n, \Gamma) > 1-\omega. \quad \ldots \quad (11.1)$$

The set of ω-cluster values will be denoted by $C((f, \Gamma), \omega)$, or by $C(\omega)$ if there is no ambiguity, and we also define $C(0) = \bigcap_{\omega>0} C(\omega)$. A point α is in $C(0)$ if and only if there is a z_n sequence as above except that the last condition is replaced by $\lim_{n\to\infty} \mu(z_n, \Gamma) = 1$. The set $C(0)$ is closed. As ω increases, $C(\omega)$ increases in the wide sense. If Γ is an arc with 1 at one endpoint, of length $< 2\pi$, $C(0)$ is the set of tangential cluster values of f at 1, on approach from the same side as Γ, and $C(\omega)$ is the set of cluster values of f at 1 on approach in an angle formed by the tangent ray to the perimeter at 1, on the same side as Γ, and a ray to 1 making any angle $< \pi\omega$ with this tangent ray. (The factor π here is not present in the corresponding discussion in Doob (1963)).

If $0 < \omega \leqslant 1$ a value α will be said to be in the ω range of the (f, Γ) pair if there is a sequence $\{z_n, n \geqslant 1\}$ for which (11.1) is true except that the second condition is replaced by $f(z_n) = \alpha$. The ω range will be denoted by $R((f, \Gamma), \omega)$, or by $R(\omega)$ if there is no ambiguity, and we define $R(0) = \bigcap_{\omega>0} R(\omega)$. Obviously $R(\omega) \subset C(\omega)$.

It is clear that B, $C(\omega)$, $R(\omega)$ depend only on the part of Γ near 1 and that, if L is a linear transformation of the unit disc onto itself with $L(1) = 1$ then $B(f, \Gamma) = B(f(L), L(\Gamma))$ and the corresponding relations for $R(\omega)$ and $C(\omega)$ are also true.

Let $\{z_n, n \geqslant 1\}$ be a sequence of points in the unit disc with limit 1. Let $A_n(d)$ be the subset of D for which the hyperbolic distance from z_n is less than d. If $0 < \omega \leqslant 1$ and if

$$\inf_n \mu(z_n, \Gamma) > 1-\omega, \lim_{d \to \infty} \inf_n \sup_{z \in A_n(d)} \mu(z, \Gamma) = 1, \qquad \ldots \quad (11.2)$$

the set Γ will be said to be weakly ω-cohesive for the z_n sequence. Let $A_n(d, \varepsilon)$ be the open component containing z_n of $A_n(d) \cap \{z : \mu(z, \Gamma) > \varepsilon\}$. The set Γ will be said to be ω-cohesive for the z_n sequence if

$$\inf_n \mu(z_n, \Gamma) > 1-\omega, \lim_{d \to \infty} \inf_n \sup_{z \in A_n(d, 1-\omega')} \mu(z, \Gamma) = 1, \qquad \ldots \quad (11.3)$$

when $\omega-\omega'$ is positive and sufficiently small.

We shall say that a pair (f, Γ) is [weakly] ω-cohesive for the cluster value α if there is a z_n sequence for which Γ is [weakly] ω-cohesive and for which $f(z_n) \to \alpha$. The definition of simultaneous [weak] ω-cohesiveness for the points of a set is then carried through as in the analogous discussion in Section 5.

If Γ is an arc of the unit disc perimeter, with endpoint 1, in the fourth quadrant, and if $z_n \to 1$ below a ray to 1 making an angle $< \pi\omega$ with the ray tangent to the perimeter at 1 and going down, the arc Γ is ω-cohesive for the z_n sequence. Thus (f, Γ) is ω-cohesive for all cluster values in $C(\omega)$.

Let $f_n = f$ for $n \geqslant 1$ and let Γ_n be the intersection of Γ with an open perimeter arc containing 1, whose length goes to 0 with $1/n$. Then $B_1^0(f, \Gamma) = B_1^0(\{(f_n, \Gamma_n)\})$ and the corresponding equality holds for the other nine types of boundary cluster sets and for $C(\omega)$, $R(\omega)$. Moreover Γ is [weakly] ω-cohesive for a sequence $\{z_n, n \geqslant 1\}$ if and only if some subsequence of the z_n sequence is [weakly] ω-cohesive for the (f_n, Γ_n) sequence.

We conclude that all the theorems of Sections 6, 7, 9 and 10 on B, $C(\omega)$ and $R(\omega)$ are true with the definitions of the present section. The individual (f, Γ) theorems obtained in this way generalize the theorems in Doob (1963) which treat the case in which Γ is an open arc abutting the point 1. As is shown there, stronger conclusions can be obtained using B_1^0 in that case; the sets of zero capacity in the individual function-set version of Theorems 10.1 and 10.2 are empty. The corresponding (f_n, Γ_n) results were obtained in Doob (1933).

REFERENCES

BRELOT, M. (1956) : Le Problème de Dirichlet. Axiomatique et frontière de Martin. *J. Math. Pures Appl.*, **35**, 297–335.

CONSTANTINESCU, C. and CORNEA, A. (1960): Über das Verhalten der analytischen Abbildungen Riemannscher Flächen auf dem idealen Rand von Martin. *Nagoya Math. J.*, **17**, 1–87.

DOOB, J. L. (1933) : The boundary values of analytic functions. II. *Trans. Amer. Math. Soc.*, **35**, 418–451.

———— (1961) : Conformally invariant cluster value theory. *Ill. J. Math.*, **5**, 521–549.

———— (1963) : One-sided cluster value theorems. *Proc. London Math. Soc.* (3) 13 to appear.

NOSHIRO, KIYOSHI (1960) : *Cluster Sets*, Springer Verlag, Berlin.

Paper received : January, 1963.

VON MISES FUNCTIONALS AND MAXIMUM LIKELIHOOD ESTIMATION*

By G. KALLIANPUR

Michigan State University

1. Introduction

In 1955, C. R. Rao and the author introduced a class of estimators of a parameter θ which were Fisher consistent (FC) and which as functionals of distribution functions, satisfied suitable regularity conditions such as Fréchet differentiability (Kallianpur and Rao, 1955). It was shown that any statistic belonging to this class is consistent and asymptotically normally distributed with asymptotic variance greater than or equal to $[n\, i(\theta)]^{-1}$, where $i(\theta)$ is Fisher's information function.

In view of this result, R. A. Fisher's definition of efficiency (for a recent discussion of this and related concepts see Rao, 1952) becomes meaningful or justifiable at least so far as the class of FC, Fréchet differentiable estimators is concerned. However, such an approach can be considered useful or interesting provided this class is general enough and if it can be shown that there are estimators belonging to this class that are efficient. In particular, it would be desirable to prove that the likelihood estimator is a member of this class (and hence efficient with respect to this class). Although, in a later paper Rao (1957) was able to show this to be the case when θ is a parameter in a multinomial distribution, the question in general, remained unanswered in the joint paper by the author and Rao (1955). The authors of that paper were not able, under any reasonable set of assumptions (on the density function in the continuous and the probability function in the infinite discrete case), to prove the Fréchet differentiability of the ML estimator. Apparently, at the root of the difficulty is the fact that Fréchet differentiability is too severe a restriction when dealing with the "infinite" dimensional (i.e. non-multinomial) situation.

In this article we propose to examine the problem afresh by replacing Fréchet differentiability by a weaker analytical concept (which, consequently yields a wider class of estimators), that of Volterra (or V) differentiability (also called weak or Gâteaux differentiability). The notion of V-differentiable functionals was introduced into statistical work by R. von Mises as early as 1948 (cf. von Mises, 1947) and was known to the authors (Kallianpur and Rao, 1955) who, however, failed to realize its suitability at the time.

We whall consider Fisher consistent, von Mises functionals of the second order whose precise definition is given in the next section. All the results corresponding to those in Kallianpur and Rao (1955) will be derived also for estimators belonging to the new class. But the main concern of the paper will be to prove that under suitable

---

*This research was partially supported by the National Science Foundation, Grant No. 18976, and the U.S. Army Research Office, Grant No. CRD-AA-L-3782, both to Michigan State University.

This paper has been included in *Contributions to Statistics* presented to Professor P. C. Mahalanobis on the occasion of his 70th birthday.

conditions, the ML estimator is a FC, von Mises functional of the second order. Thus the class of FC, von Mises functionals of the second order possesses desirable features mentioned earlier, which the class of FC, Fréchet differentiable functionals, has not been shown to have. Although the conditions under which the V-differentiability and efficiency of the ML estimator are proved are more stringent than the conditions needed for showing its consistency and asymptotic normality, they are of the same general nature and apply to a reasonably wide class of distributions. Hence no attempt has been made to seek refinements in this direction. The interested reader will, no doubt, be able to improve upon them.

Finally, the closing section offers some remarks on another concept of efficiency recently introduced by Rao (1962).

2. von Mises functionals of the second order

The following notation will be adhered to throughout what follows :

(i) The parameter to be estimated, θ, lies in an open interval I of the real line;

(ii) F_θ stands for the common distribution function of the independent observations $X_1, \ldots, X_n$ when θ is the "true" value of the parameter;

(iii) P_θ is the infinite product measure determined by F_θ on the (infinite) product of the real line with itself, a generic point of the latter space being denoted by $\omega = (x_1, x_2, \ldots)$.

(iv) For each n, and $\omega = (x_1, x_2, \ldots)$, $F_n^*(\cdot, \omega)$ denotes the empirical distribution function when the observed sample is given by $(x_1, \ldots, x_n)$.

The definition of a von Mises functional of the second order given below is a slight modification of the one adopted in a recent paper by Filippova (1962).

We shall consider functionals T defined on a subset S_T of the space of all distribution functions (d.f.'s). It will be assumed that F_θ belongs to S_T for all θ in I. A subset τ_T of S_T is said to be starlike at F_θ if for every $W \epsilon \tau_T$ and $t \epsilon [0, 1]$, the d.f. $F_\theta + t(W - F_\theta) \epsilon \tau_T$.

A functional T is m times differentiable at F_θ in the Volterra sense (or m times V-differentiable) relative to τ_T, starlike at F_θ if for $k = 1, \ldots, m$.

(a) $\frac{d^k}{dt^k} T[F_\theta + t(W - F_\theta)]$ exists for every t in $[0, 1]$ and every W in τ_T, and

(b) there exists a functional $T^{(k)}[F_\theta; x_1, \ldots, x_k]$ depending on F_θ and k real variables $x_1, \ldots, x_k$ such that for every $W \epsilon \tau_T$, (writing $h = W - F_\theta$)

$$\frac{d^k}{dt^k} T[F_\theta + th]_{t=0} = \int \ldots \int T^{(k)}[F_\theta; x_1, \ldots, x_k] \, dh(x_1) \cdots dh(x_k).$$

Here, as in the sequel, all integrations are over the entire domain of definition of the variables. The functional $T^{(k)}[F_\theta; x_1, \ldots, x_k]$ is assumed to be a Borel function of $(x_1, \ldots, x_k)$ and is called the k-th V-derivative of T at F_θ.

We are now in a position to introduce our basic definition.

Definition 1 : T is a von Mises functional of the second order at F_θ if the following three conditions are fulfilled :

(i) There exists a set $\tau_T(\theta)$ starlike at F_θ such that

$$\lim_{n\to\infty} P_\theta(\omega : F_n^*(\cdot, \omega)\epsilon\tau_T(\theta)] = 1;$$

(ii) T is twice V-differentiable at F_θ relative to $\tau_T(\theta)$;

(iii) for every $\epsilon > 0$,

$$\lim_{n\to\infty} P_\theta^*\{\omega : n^{\frac{1}{2}}d_n(\omega) \geqslant \epsilon\} = 0,$$

where

$$d_n(\omega) = \sup_{0\leqslant t\leqslant 1}\left|\frac{d^2}{dt^2}\,T[F_n^{(t)}(\cdot,\,\omega)]\right|,$$

$$F_n^{(t)}(\cdot,\,\omega) = F_\theta(\cdot)+t\{F_n^*(\cdot,\,\omega)-F_\theta(\cdot)\},$$

and P_θ^* is the outer measure corresponding to P_θ.

Condition (iii) of the definition is stated in terms of outer measure since this is all that is needed for our purpose and to avoid any discussion of the measurability $d_n(\omega)$. Observe that $d_n(\omega)$ is assumed and defined only for ω in the set $M_{n,T} = (\omega : F_n(\cdot,\,\omega)\epsilon\tau_T(\theta))$ so that the set displayed in (ii) is necessarily a subset of $M_{n,T}$. Although $T[F_n(\cdot,\,\omega)]$ is defined only for $\omega\epsilon M_{n,T}$ its definition can be formally completed for all ω in many ways (for instance, by setting it equal to a constant when $\omega \notin M_{n,T}$. Since $P_\theta(M'_{n,T})$ tends to zero, the values of $T[F_n(\cdot,\,\omega)]$ on $M_{n,T}$ do not in any way affect the asymptotic results in which we are interested. It will be assumed that the $T[F_n(\cdot,\,\omega)]$ thus obtained is, for each n, a measurable ω -function.

(iv) If $F_{\theta+\eta}\,\epsilon\tau_T(\theta)$ for all η sufficiently small (θ being the true value of the parameter) then

$$\sup_{0\leqslant t\leqslant 1}\left|\frac{d^2}{dt^2}\;T[F_\theta+t(F_{\theta+\eta}-F_\theta)]\right| = 0(\eta) \text{ as } \eta\to 0.$$

The conditions to be imposed on F_θ will ensure that the first part of (iv) is always fulfilled so that essentially (iv) will be a restriction on the functional T.

We shall suppose that the random variables have a common probability density $f(x,\,\theta)$ with respect to some fixed σ-finite measure λ. The regularity conditions under which our results will be obtained are naturally stronger than those needed to prove only consistency and asymptotic normality. The following assumptions will be assumed to hold throughout the rest of this paper.

(A) These are essentially the conditions assumed by Cramér (1956). For every x except at most on a set of λ measure zero, the derivatives $\frac{\partial^i \log f}{\partial\theta^i}$ $(i = 1,\,2,\,3)$ exist for every θ in I. For any $\theta\,\epsilon\,I\left|\frac{\partial^i f}{\partial\theta^i}\right| \leqslant G_i$ where G_i are integrable with respect to λ over $(-\infty,\,\infty)$, while $\left|\frac{\partial^3 \log f}{\partial\theta^3}\right| \leqslant H$ for all θ in I where $E_\theta H \leqslant K < \infty$, K being independent of θ.

For every θ in I, $0 < E_\theta\left(\frac{\partial \log f}{\partial\theta}\right)^2 < \infty.$

(B) Let $a_0(x, \theta) = \log f(x, \theta)$, $a_i(x, \theta) = \dfrac{\partial^i \log f(x, \theta)}{\partial \theta^i}$ $(i = 1, 2)$ and $h(\theta) = E_\theta H$.

(i) For each θ in I there exists a neighbourhood N of θ such that $E_{\theta'}[a_i(X, \theta)]$ $(i = 0, 1, 2)$ exists for $\theta' \epsilon N$.

(ii) For $i = 1, 2$, $E_\theta[a_i^2(X, \theta)] < \infty$ and $E_\theta H^2 < \infty$ for each θ in I.

(iii) For every θ in I, the functions $g_\eta(x, \theta) = (\eta f(x, \theta))^{-2}[f(x, \theta+\eta) - f(x, \theta)]^2$ (η being positive and tending to zero) are uniformly integrable with respect to F_θ.

Conditions (ii) and (iii) of (B) are somewhat stronger than are actually needed in the proof of the results. A consequence of (ii) and (iii) is that the functions $E_{\theta'}[a_i(X, \theta)]$ $(i = 1, 2,)$ are continuous functions of θ' in N, differentiable with respect to θ' at $\theta' = \theta$ with derivative equal to $E_\theta[a_i(X, \theta)\ a_1(X, \theta)]$.

We recall that a functional T is Fisher consistent if $T[F_\theta] = \theta$ for all θ in I.

Now if T is any FC, von Mises functional of the second order (we shall omit the phrase "at F_θ"), it is easy to verify that

$$\begin{aligned} n^{\frac{1}{2}} | \{T[F_n(\cdot, \omega)] - \theta\} - n^{\frac{1}{2}} \int T^{(1)}[F_\theta; x]\, d[F_n(\cdot, \omega) - F_\theta] | \\ \leqslant n^{\frac{1}{2}} d_n(\omega)\, c_{M_{n,T}}(\omega) + \xi_n(\omega), \end{aligned}$$

where $c_{M_{n,T}}$ is the characteristic function of $M_{n,T}$ and ξ_n converges to zero in probability. From condition (iii) of Definition 1, it follows that the random variable on the left side of the above inequality converges to zero in probability. From this, and a well known result of Khinchin (see Kallianpur and Rao, 1955) on the normal convergence of sums of independent, identically distributed random variables we immediately have our first result.

Theorem 1 : *If T is a FC, von Mises functional of the second order, then $n^{\frac{1}{2}}\{T[F_n] - \theta\}$ is asymptotically normally distributed if and only if $\int (T^{(1)}[F_\theta; x])^2\, dF_\theta(x)$ is finite.*

Theorem 1 is the analogue of Theorem 3 of Kallianpur and Rao (1955) stated there for Fréchet differentiable FC functionals. Let $\mathscr{M}$ be the class of all measurable, FC, von Mises functionals of the second order for which $\int (T^{(1)}[F_{\theta_0}; x])^2\, dF_{\theta_0}(x)$ is finite. By imposing suitable regularity conditions on F_θ a lower bound for the asymptotic variance of $T \epsilon \mathscr{M}$ will be obtained, as in Kallianpur and Rao (1955), which coincides with the bound given by Fisher.

3. The maximum likelihood functional and its properties

Now let θ_0 denote the unknown true value of θ and for δ a sufficiently small number to be specified later let U_δ be the open interval $(\theta_0 - \delta, \theta_0 + \delta)$ whose closure $(\bar{U}_\delta)$ is contained in I. Let $\Lambda_\delta(\theta_0)$ be the set of all distribution functions W satisfying the following conditions :

(i) The integrals $\int a_i(x, \theta) dW$ $(i = 0, 1, 2)$ exist and are continuous functions of θ in $\bar{U}_\delta$;

(ii)
$$| \int a_1(x, \theta_0) dW | < \delta^2, \int a_2(x, \theta_0) dW < -\tfrac{1}{2} k^2$$

where we write

$$k^2 = E_{\theta_0}\left(\frac{\partial \log f}{\partial \theta}\right)_0^2, \text{ and } \int H(x)dW < 2K.$$

The set $\Lambda_\delta(\theta_0)$ is clearly non empty since it contains F_{θ_0}. In view of conditions (B) it is easily seen that F_θ belongs to $\Lambda_\delta(\theta_0)$ for all θ in some $N(\theta_0)$. This fact will be used later. Now define $S_\delta(\theta_0)$ to be the set of all distributions $V^{(t)} = F_{\theta_0}+t(W-F_{\theta_0})$, where $W \epsilon \Lambda_\delta(\theta_0)$ and $0 \leqslant t \leqslant 1$. It is easy to verify that $S_\delta(\theta_0)$ is starlike at F_{θ_0}. For any V in $S_\delta(\theta_0)$, $\phi(\theta, V) = \int a_0(x, \theta)dV$ exists and defines, for every θ in $\bar{U}_\delta$, a functional over $S_\delta(\theta_0)$. By the definition of $\Lambda_\delta(\theta_0)$, $\phi'(\theta, V)$ (primes denote differentiations with respect to θ) exists and is given by $\int a_1(x, \theta)dV$ which is continuous in $\bar{U}_\delta$. Taking $V = V_t$ and applying the finite Taylor expansion to $\phi'(\theta, V^{(t)})$ we write

$$\begin{aligned}\phi'(\theta, V^{(t)}) &= \phi'(\theta_0, V^{(t)})+(\theta-\theta_0)\int a_2(x, \theta_0)dV^{(t)}+\tfrac{1}{2}\beta(\theta-\theta_0)^2\, HdV^{(t)} \\ &= -k^2(\theta-\theta_0)+A_{t'},\end{aligned} \qquad \ldots (3.1)$$

where $|\beta| \leqslant 1$ and $|A_t| \leqslant \delta^2(2K+1)+\frac{1}{2}k^2\delta$. From this it is easy to see that $\phi'(\theta_0-\delta, V^{(t)}) > 0$ and $\phi'(\theta_0+\delta, V^{(t)}) < 0$ provided $0 < \delta < \frac{1}{2}k^2(3K+1)^{-1}$. Since $\phi'(\theta, V^{(t)})$ is continuous in $\bar{U}_\delta$, it follows that for each $V^{(t)}$ in $S_\delta(\theta_0)$ the equation $\phi'(\theta, V^{(t)}) = 0$ has a root in U_δ. This root, which we shall denote by $\hat{\theta}[V^{(t)}]$, is a function or functional of $V^{(t)}$ and is defined on $S_\delta(\theta_0)$. Furthermore, again utilizing the definition of $\Lambda_\delta(\theta_0)$ and $V^{(t)}$, we have $\phi''(\theta, V^{(t)})) \leqslant -\frac{1}{2}k^2+4K\delta < 0$ for all θ in U_δ, δ being chosen such that $0 < \delta < \min\{\frac{1}{2}k^2(3K+1)^{-1}, \frac{1}{8}k^2K^{-1}\}$. It follows that $\hat{\theta}[V^{(t)}]$ is the unique root of the equation $\phi'(\theta, V^{(t)}) = 0$ in U_δ as also the unique maximum of $\phi(\theta\ V^{(t)})$ in U_δ.

Essentially the same argument as Cramér's (1946) consistency proof shows that the P_{θ_0}-probability that $F_n^*(\cdot, \omega)$ belongs to $\Lambda_\delta(\theta_0)$ tends to one as $n \to \infty$ and hence that for all t in $[0, 1]P_{\theta_0}[\omega\colon F_n^{(t)}(\cdot, \omega)\epsilon S_\delta(\theta_0)] \to 1$. Observe that whenever $F_n^*(\cdot, \omega)\epsilon\, S_\delta(\theta_0)$, the likelihood function of the sample coincides on U_δ with the function $\phi(\theta, F_n^*)$. It follows that the unique root of the likelihood equation lying in U_δ for which the likelihood is a local maximum (the existence of such a root is a consequence of a combination of arguments of Cramér (1946) and Huzurbazar (1948) and need not be repeated here) is indeed given by $\hat{\theta}[F_n^*(\cdot, \omega)]$ where $\hat{\theta}$ is the functional defined above. We shall refer to $\hat{\theta}$ as the *maximum likelihood functional* (MLF) although, as is well known, $\hat{\theta}[F_n^*(\cdot, \omega)]$ does not necessarily make the likelihood an absolute maximum. In accordance with the notation of the preceding section we shall write $\tau_{\hat{\theta}}(\theta_0)$ instead of $S_\delta(\theta_0)$.

The proof that the MLF $\hat{\theta}$ is a von Mises functional of the second order consists of two parts. First we show its Fisher consistency and derive some of its analytical properties regarded as a functional defined on $\tau_{\hat{\theta}}(\theta_0)$.

Theorem 2 : *The ML functional $\hat{\theta}$ is FC and is twice V-differentiable relative to the set $\tau_{\hat{\theta}}(\theta_0)$. The first and second order V-derivatives of $\hat{\theta}$ at F_{θ_0} are given by $k^{-2}a_1(x, \theta_0)$ and $2k^{-4}a_1(x_1, \theta_0)\,a_2(x_2, \theta_0)$ respectively.*

(B) Let $a_0(x, \theta) = \log f(x, \theta)$, $a_i(x, \theta) = \dfrac{\partial^i \log f(x, \theta)}{\partial \theta^i}$ $(i = 1, 2)$ and $h(\theta) = E_\theta H$.

(i) For each θ in I there exists a neighbourhood N of θ such that $E_{\theta'}[a_i(X, \theta)]$ $(i = 0, 1, 2)$ exists for $\theta' \epsilon N$.

(ii) For $i = 1, 2$, $E_\theta[a_i^2(X, \theta)] < \infty$ and $E_\theta H^2 < \infty$ for each θ in I.

(iii) For every θ in I, the functions $g_\eta(x, \theta) = (\eta f(x, \theta))^{-2}[f(x, \theta+\eta)-f(x, \theta)]^2$ (η being positive and tending to zero) are uniformly integrable with respect to F_θ.

Conditions (ii) and (iii) of (B) are somewhat stronger than are actually needed in the proof of the results. A consequence of (ii) and (iii) is that the functions $E_{\theta'}[a_i(X, \theta)]$ $(i = 1, 2,)$ are continuous functions of θ' in N, differentiable with respect to θ' at $\theta' = \theta$ with derivative equal to $E_\theta[a_i(X, \theta)\ a_1(X, \theta)]$.

We recall that a functional T is Fisher consistent if $T[F_\theta] = \theta$ for all θ in I.

Now if T is any FC, von Mises functional of the second order (we shall omit the phrase "at F_θ"), it is easy to verify that

$$n^{\frac{1}{2}}|\{T[F_n(\cdot, \omega)]-\theta\}-n^{\frac{1}{2}}\int T^{(1)}[F_\theta; x]\, d[F_n(\cdot, \omega)-F_\theta]| \leqslant n^{\frac{1}{2}}d_n(\omega)\, c_{M_{n,T}}(\omega)+\xi_n(\omega),$$

where $c_{M_{n,T}}$ is the characteristic function of $M_{n,T}$ and ξ_n converges to zero in probability. From condition (iii) of Definition 1, it follows that the random variable on the left side of the above inequality converges to zero in probability. From this, and a well known result of Khinchin (see Kallianpur and Rao, 1955) on the normal convergence of sums of independent, identically distributed random variables we immediately have our first result.

Theorem 1 : *If T is a FC, von Mises functional of the second order, then $n^{\frac{1}{2}}\{T[F_n]-\theta\}$ is asymptotically normally distributed if and only if $\int (T^{(1)}[F_\theta; x])^2\, dF_\theta(x)$ is finite.*

Theorem 1 is the analogue of Theorem 3 of Kallianpur and Rao (1955) stated there for Fréchet differentiable FC functionals. Let $\mathcal{M}$ be the class of all measurable, FC, von Mises functionals of the second order for which $\int (T^{(1)}[F_{\theta_0}; x])^2\, dF_{\theta_0}(x)$ is finite. By imposing suitable regularity conditions on F_θ a lower bound for the asymptotic variance of $T \epsilon \mathcal{M}$ will be obtained, as in Kallianpur and Rao (1955), which coincides with the bound given by Fisher.

3. The maximum likelihood functional and its properties

Now let θ_0 denote the unknown true value of θ and for δ a sufficiently small number to be specified later let U_δ be the open interval $(\theta_0-\delta, \theta_0+\delta)$ whose closure $(\bar{U}_\delta)$ is contained in I. Let $\Lambda_\delta(\theta_0)$ be the set of all distribution functions W satisfying the following conditions :

(i) The integrals $\int a_i(x, \theta) dW$ $(i = 0, 1, 2)$ exist and are continuous functions of θ in $\bar{U}_\delta$;

(ii)
$$|\int a_1(x, \theta_0) dW| < \delta^2, \int a_2(x, \theta_0) dW < -\tfrac{1}{2}k^2$$

where we write

$$k^2 = E_{\theta_0}\left(\frac{\partial \log f}{\partial \theta}\right)_0^2, \text{ and } \int H(x)dW < 2K.$$

The set $\Lambda_\delta(\theta_0)$ is clearly non empty since it contains F_{θ_0}. In view of conditions (B) it is easily seen that F_θ belongs to $\Lambda_\delta(\theta_0)$ for all θ in some $N(\theta_0)$. This fact will be used later. Now define $S_\delta(\theta_0)$ to be the set of all distributions $V^{(t)} = F_{\theta_0}+t(W-F_{\theta_0})$, where $W \epsilon \Lambda_\delta(\theta_0)$ and $0 \leqslant t \leqslant 1$. It is easy to verify that $S_\delta(\theta_0)$ is starlike at F_{θ_0}. For any V in $S_\delta(\theta_0)$, $\phi(\theta, V) = \int a_0(x, \theta)dV$ exists and defines, for every θ in $\bar{U}_\delta$, a functional over $S_\delta(\theta_0)$. By the definition of $\Lambda_\delta(\theta_0)$, $\phi'(\theta, V)$ (primes denote differentiations with respect to θ) exists and is given by $\int a_1(x, \theta)dV$ which is continuous in $\bar{U}_\delta$. Taking $V = V_t$ and applying the finite Taylor expansion to $\phi'(\theta, V^{(t)})$ we write

$$\begin{aligned}\phi'(\theta, V^{(t)}) &= \phi'(\theta_0, V^{(t)})+(\theta-\theta_0)\int a_2(x, \theta_0)dV^{(t)}+\tfrac{1}{2}\beta(\theta-\theta_0)^2 \, HdV^{(t)} \\ &= -k^2(\theta-\theta_0)+A_{t'},\end{aligned} \qquad \dots \quad (3.1)$$

where $|\beta| \leqslant 1$ and $|A_t| \leqslant \delta^2(2K+1)+\frac{1}{2}k^2\delta$. From this it is easy to see that $\phi'(\theta_0-\delta, V^{(t)}) > 0$ and $\phi'(\theta_0+\delta, V^{(t)}) < 0$ provided $0 < \delta < \frac{1}{2}k^2(3K+1)^{-1}$. Since $\phi'(\theta, V^{(t)})$ is continuous in $\bar{U}_\delta$, it follows that for each $V^{(t)}$ in $S_\delta(\theta_0)$ the equation $\phi'(\theta, V^{(t)}) = 0$ has a root in U_δ. This root, which we shall denote by $\hat{\theta}[V^{(t)}]$, is a function or functional of $V^{(t)}$ and is defined on $S_\delta(\theta_0)$. Furthermore, again utilizing the definition of $\Lambda_\delta(\theta_0)$ and $V^{(t)}$, we have $\phi''(\theta, V^{(t)})) \leqslant -\frac{1}{2}k^2 + 4K\delta < 0$ for all θ in U_δ, δ being chosen such that $0 < \delta < \min\{\frac{1}{2}k^2(3K+1)^{-1}, \frac{1}{8}k^2K^{-1}\}$. It follows that $\hat{\theta}[V^{(t)}]$ is the unique root of the equation $\phi'(\theta, V^{(t)}) = 0$ in U_δ as also the unique maximum of $\phi(\theta\ V^{(t)})$ in U_δ.

Essentially the same argument as Cramér's (1946) consistency proof shows that the P_{θ_0}-probability that $F_n^*(\cdot, \omega)$ belongs to $\Lambda_\delta(\theta_0)$ tends to one as $n \to \infty$ and hence that for all t in $[0, 1] P_{\theta_0}[\omega\colon F_n^{(t)}(\cdot, \omega) \epsilon S_\delta(\theta_0)] \to 1$. Observe that whenever $F_n^*(\cdot, \omega) \epsilon S_\delta(\theta_0)$, the likelihood function of the sample coincides on U_δ with the function $\phi(\theta, F_n^*)$. It follows that the unique root of the likelihood equation lying in U_δ for which the likelihood is a local maximum (the existence of such a root is a consequence of a combination of arguments of Cramér (1946) and Huzurbazar (1948) and need not be repeated here) is indeed given by $\hat{\theta}[F_n^*(\cdot, \omega)]$ where $\hat{\theta}$ is the functional defined above. We shall refer to $\hat{\theta}$ as the *maximum likelihood functional* (MLF) although, as is well known, $\hat{\theta}[F_n^*(\cdot, \omega)]$ does not necessarily make the likelihood an absolute maximum. In accordance with the notation of the preceding section we shall write $\tau_{\hat{\theta}}(\theta_0)$ instead of $S_\delta(\theta_0)$.

The proof that the MLF $\hat{\theta}$ is a von Mises functional of the second order consists of two parts. First we show its Fisher consistency and derive some of its analytical properties regarded as a functional defined on $\tau_{\hat{\theta}}(\theta_0)$.

Theorem 2 : *The ML functional $\hat{\theta}$ is FC and is twice V-differentiable relative to the set $\tau_{\hat{\theta}}(\theta_0)$. The first and second order V-derivatives of $\hat{\theta}$ at F_{θ_0} are given by $k^{-2}\, a_1(x, \theta_0)$ and $2k^{-4}a_1(x_1, \theta_0)\, a_2(x_2, \theta_0)$ respectively.*

Proof : Setting $t = 0$ and $\theta = \hat{\theta}$ in equation (3.1)

we have $$[\hat{\theta}[F_{\theta_0}]-\theta_0][-k^2+\tfrac{1}{2}\beta\{\hat{\theta}[F_{\theta_0}]-\theta_0\}\textstyle\int HdF_{\theta_0}] = 0.$$

Since $|\hat{\theta}[F_{\theta_\theta}]-\theta_0| < \delta$ the second factor on the right being less than $-k^2+K\delta$ cannot be zero so that $\hat{\theta}[F_{\theta_0}] = \theta_0$. This proves Fisher consistency. First we show that $\hat{\theta}[V^{(t)}]$ is a continuous function of t in [0, 1]. In subsequent arguments we shall use the expression "for δ sufficiently small" to mean that δ has been chosen less than a fixed positive number δ_0 which depends on k^2 and K. That δ can be so determined will be apparent from the context. Now writing $V^{(t)} = F_{\theta_0}+th$ $(h = W-F_{\theta_0})$, a_i for $a_i(x, \theta_0)$ $(i = 1, 2)$ and letting t' be such that $0 \leqslant t+t' \leqslant 1$ it follows from (3.1) after some simplification that

$$\begin{aligned}|\hat{\theta}[V^{(t+t')}]-\hat{\theta}[V^{(t')}]| \leqslant |t'|[|\textstyle\int a_2dh \int a_1dV^{(t)}-\int a_1dh\int a_2dV^{(t)}| \\ +\tfrac{1}{2}(|\textstyle\int a_1dV^{(t)} \int Hdh|+|\int HdV^{(t)} \int a_1dh|)] \\ \times[(k^2-(t+t')\textstyle\int a_2dh-\tfrac{1}{2}\delta(|\int HdV^{(t)} +t'\int Hdh|)) \\ \times(k^2-t\textstyle\int a_2dh-\tfrac{1}{2}\delta|\int HdV^{(t)}|)-\tfrac{1}{2}|\int a_1dV^{(t)} \int HdV^{(t)}|]^{-1}\end{aligned}$$

the quantity on the right side being positive for δ sufficiently small, and tending to zero as $t' \to 0$. The assertion is thus proved. Again from (3.1) we have

$$\begin{aligned}t'^{-1}\{\theta[V^{(t+t')}]-\theta[^{(t)}]\} = [\textstyle\int a_1dV^{(t)}\int a_2dh-\int a_1dh\int a_2dV^{(t)} \\ +\tfrac{1}{2}\beta\Delta(t+t')\textstyle\int a_1dV^{(t)}\int Hdh-\tfrac{1}{2}\beta\Delta(t)\int a_1dh\int HdV^{(t)}] \\ \times[\{\textstyle a_2dV^{(t)}+t'\int a_2dh+\tfrac{1}{2}\beta\Delta(t+t')(\int Hd\,V^{(t)}+t'\int H\,dh\} \\ \times\{\textstyle\int a_2dV^{(t)}+\tfrac{1}{2}\beta\Delta(t)\int HdV^{(t)}\}-\tfrac{1}{2}\beta\int a_1dV^{(t)}\int HdV^{(t)}],\end{aligned}$$

where $\Delta(t) = \hat{\theta}[V^{(t)}]-\theta_0$. Since $\Delta(t)$ has been shown to be continuous it follows on making $t' \to 0$ that $\hat{\theta}[V^{(t)}]$ is differentiable with respect to t and that

$$\frac{d}{dt}\hat{\theta}[V^{(t)}] = \frac{P(t)}{R(t)} \qquad \ldots \quad (3.2)$$

where

$$P(t) = \textstyle\int a_1dV^{(t)}\int a_2dh-\int a_1dh\int a_2dV^{(t)}+\tfrac{1}{2}\beta\Delta(t)[\int a_1dV^{(t)}\int Hdh-\int a_1dh\int HdV^{(t)}]$$

and

$$R(t) = (\textstyle\int a_2dV^{(t)}+\tfrac{1}{2}\beta\Delta(t)\int HdV^{(t)})^2-\tfrac{1}{2}\beta\int a_1dV^{(t)}\int HdV^{(t)}.$$

Since $\hat{\theta}[V^{(t)}]$ and hence $\Delta(t)$ is differentiable with respect to t, so are $P(t)$ and $R(t)$ and it follows from (3.2) that $\frac{d^2}{dt^2}\hat{\theta}[V^{(t)}]$ exists. We have (primes now denoting differentiations with respect to t)

$$\frac{d^2}{dt^2}\hat{\theta}[V^{(t)}] = \frac{P'(t)}{R(t)}-\frac{P(t)R'(t)}{R^2(t)}. \qquad \ldots \quad (3.3)$$

Finally putting $t = 0$ in (3.2) and (3.3) we obtain

$$\left(\frac{d}{dt}\hat{\theta}[V^{(t)}]\right)_{t=0} = \textstyle\int k^{-2}a_1dh,$$

and $$\left(\frac{d^2}{dt^2}\hat{\theta}[V^{(t)}]\right)_{t=0} = 2\textstyle\int\int k^{-4}a_1(x_1, \theta_0)\, a_2(x_2, \theta_0)\, dh(x_1)\, dh(x_2).$$

The proof of Theorem 2 is thus complete.

It has already been shown above that $P_{\theta_0}(M_{n,\hat{\theta}}) \to 1$ as $n \to \infty$ where $M_{n,\hat{\theta}}$ is the set $\{\omega : F_n^*(\cdot, \omega) \in \tau_{\hat{\theta}}(\theta_0)\}$. This is condition (i) of Definition 1 with $T = \hat{\theta}$. We turn now to the proof that condition (iii) of that definition holds for $\hat{\theta}$. For every positive ϵ there is an integer $n_0(\varepsilon, \delta)$ such that the following holds: $P_{\theta_0}(M_{n,\hat{\theta}}) > 1-\varepsilon$ if $n \geqslant n_0$, and whenever $\omega \epsilon M_{n,\hat{\theta}}$, for every fixed t in $[0, 1]$, $\hat{\theta}[F_n^{(t)}(\cdot, \omega)]$ is the unique root of $\phi'(\theta, F^{(t)}) = 0$ which maximizes $\phi(\theta, F_n^{(t)})$ in $\bar{U}_\delta$. The definition of $\hat{\theta}$ for all ω can be completed in such a way that $\hat{\theta}[F_n^{(t)}(\cdot, \omega)]$ is, for each t, measurable as a function of ω. To prove (iii) it suffices to assume that $n \geqslant n_0$ and that ω belongs to $M_{n,\hat{\theta}}$. Since $F_n^{(t)}(\cdot, \omega) \in \tau_{\hat{\theta}}(\theta_0)$ it then follows from Theorem 2 that $\frac{d^2}{dt^2}\hat{\theta}[F_n^{(t)}(\cdot, \omega)]$ exists and is given by $P_n' R_n^{-1} - P_n R_n' R_n^{-2}$ where P_n and R_n are the functions P and R given immediately after formula (3.2) except that now $h = F_n^*(\cdot, \omega) - F_{\theta_0}$. Hence it is to be noted that P_n and R_n as well as their derivatives with respect to t are functions of t and ω.

Let us write

$$A_n = \int a_1 dF_n^*, \quad B_n = \int a_2 dF_n^*, \quad C_n = \int H\, dF_n^*,$$

$$\Delta_n(t, \omega) = \hat{\theta}[F_n^{(t)}(\cdot, \omega)] - \theta_0 \quad \text{and} \quad \Delta_n(\omega) = \sup_{0 \leqslant t \leqslant 1} |\Delta_n(t, \omega)|.$$

To simplify the writing we shall suppress ω throughout the ensuing argument, but it will be supposed that $\omega \in M_{n,\hat{\theta}}$, and that all sets considered are ω-sets. Cleraly $\Delta_n \leqslant \delta$ and further from (3.1) (writing $\Delta_n(t)$ for $\Delta_n(t, \omega)$),

$$\Delta_n(t) = tA_n \cdot [k^2 - t(B_n + k^2) - \tfrac{1}{2}\beta\Delta_n(t)\{tC_n + (1-t)\int HdF_{\theta_0}\}]^{-1}.$$

Hence
$$\Delta_n \leqslant |A_n| \cdot [k^2 - |B_n + k^2| - 2K\delta]^{-1} = \xi_n, \text{ say.} \qquad \ldots (3.4)$$

Clearly ξ_n converges to zero in probability by Slutsky's theorem since by the law of large numbers A_n and $B_n + k^2$ both converge to zero in probability and since δ is sufficiently small. Hence for every $\varepsilon' > 0$, we have

$$P^*_{\theta_0}[(\Delta_n > \varepsilon') \cap M_{n,\hat{\theta}}] \leqslant P_{\theta_0}(\xi_n > \varepsilon') \to 0. \qquad \ldots (3.5)$$

Now, from the expressions for $P_n(t)$ and $R_n(t)$ we have for any t in $[0, 1]$

$$|R_n(t)| \geqslant [k^2 - |B_n + k^2| - \tfrac{1}{2}\Delta_n\{C_n + \int HdF_{\theta_0}\}]^2 - \tfrac{1}{2}|A_n|\{C_n + \int HdF_{\theta_0}\}$$

$$\geqslant (k^2 - |B_n + k^2| - 2K\delta)^2 - 2K|A_n| = \rho_n, \text{ say,}$$

while
$$|P_n(t)| \leqslant |A_n|[k^2 + K\delta].$$

Hence
$$\sup_{0 \leqslant t \leqslant 1} \left| \frac{P_n(t)}{R_n(t)} \right| \leqslant |A_n|[k^2 + K\delta]\rho_n^{-1}. \qquad \ldots (3.6)$$

We then have for $\gamma\left(\text{here and in the sequel } \gamma \text{ is a positive number less than } \frac{1}{2}\right)$

$$P^*_{\theta_0}\left(\left[n^{\frac{1}{2}-\gamma} \sup_{0\leqslant t\leqslant 1}\left|\frac{P_n(t)}{R_n(t)}\right| > \varepsilon'\right] \cap M_{n,\hat{\theta}}\right) \leqslant P_{\theta_0}[n^{\frac{1}{2}-\gamma}|A_n|[k^2+K\delta]\rho_n^{-1} > \varepsilon']$$

which tends to zero as $n\to\infty$ again by Slutsky's theorem since $\rho_n \to (k^2-2K\delta)^2\ (>0)$ in probability and $n^{\frac{1}{2}} A_n$ has an asymptotic normal distribution. Finally since

$$\sup_{0\leqslant t\leqslant 1}\left|\frac{d^2}{dt^2}\hat{\theta}[F_n^{(t)}]\right| \leqslant \sup_{0\leqslant t\leqslant 1}\left|\frac{P_n'(t)}{R_n(t)}\right| + \sup_{0\leqslant t\leqslant 1}\left|\frac{P_n(t)}{R_n(t)}\frac{R_n'(t)}{R_n(t)}\right|$$

we have

$$P^*_{\theta_0}\left\{\left(n^{\frac{1}{2}} \sup_{0\leqslant t\leqslant 1}\left|\frac{d^2}{dt^2}\hat{\theta}[F_n^{(t)}]\right| > \varepsilon'\right)\cap M_{n,\hat{\theta}}\right\}$$

$$\leqslant P^*_{\theta_0}\left\{\left(n^{\frac{1}{2}} \sup_{0\leqslant t\leqslant 1}\left|\frac{P_n'(t)}{R_n(t)}\right| > \frac{1}{2}\varepsilon'\right)\cap M_{n,\hat{\theta}}\right\}$$

$$+P^*_{\theta_0}\left\{\left(n^{\frac{1}{2}} \sup\left|\frac{P_n(t)}{Rn(t)}\cdot\frac{R_n'(t)}{R_n(t)}\right| > \frac{1}{2}\varepsilon'\right)\cap M_{n,\hat{\theta}}\right\} \quad \dots \quad (3.7)$$

It is easily seen that

$$P_n'(t) = -\left(\frac{1}{2}\beta A_n \int H dF_{\theta_0}\right)\cdot\frac{P_n(t)}{R_n(t)}.$$

From this we obtain $\quad \sup|P_n'(t)| \leqslant K|A_n|\sup\left|\frac{P_n(t)}{R_n(t)}\right|.$

Hence
$$n^{\frac{1}{2}}\sup\left|\frac{P_n'(t)}{R_n(t)}\right| \leqslant K(n^{\gamma}|A_n|\rho_n^{-1})\left(n^{\frac{1}{2}-\gamma}\sup\left|\frac{P_n(t)}{R_n(t)}\right|\right)$$

from which it follows that the first term of the right side of inequality (3.7) tends to zero. Using (3.5) we obtain after straightforward calculations that

$$\sup\left|\frac{R_n'(t)}{R_n(t)}\right| \leqslant \Big[2(k^2+|B_n+k^2|+3K\xi_n)(|B_n+k^2|$$

$$+2K\xi_n+3K\sup\left|\frac{P_n(t)}{R_n(t)}\right|+5K|A_n|\Big]\rho^{-1}.$$

Since $n^{\gamma}|B_n+k^2|$, $n^{\gamma}\xi_n$ and $n^{\gamma}|A_n|$ converge to zero in probability $(0<\gamma<\frac{1}{2})$ it follows from the above inequality and (3.7) that

$$P^*_{\theta_0}\left[\left(n^{\gamma}\sup\left|\frac{R_n'(t)}{R_n(t)}\right| > \varepsilon'\right)\cap M_{n,\hat{\theta}}\right] \to 0$$

for every $\varepsilon' > 0$.

The second term on the right side of (3.7) is not greater than

$$P^*_{\theta_0}\left[\left(n^{\gamma}\sup\left|\frac{R_n'(t)}{R_n(t)}\right|>\sqrt{\varepsilon'}\right)\cap M_{n,\hat\theta}\right]+P^*_{\theta_0}\left[n^{\frac{1}{2}-\gamma}\sup\left|\frac{P_n(t)}{R_n(t)}\right|>\frac{1}{2}\sqrt{\varepsilon'}\right)\cap M_{n,\hat\theta}\right].$$

Since both terms in the above expression approach zero as $n\to\infty$ it follows that the quantity on the left side of inequality (3.7) tends to zero as $n\to\infty$.

This completes the proof that $\hat\theta$ satisfies condition (iii) of Definition 1. The verification of condition (iv) follows without difficulty from (B). Since condition (i) has already been shown to hold these facts combined with Theorem 2 give us our main result concerning $\hat\theta$.

Theorem 3 : *The ML functional $\hat\theta$ is a von Mises functional of the second order.*

4. Efficiency of θ relative to the class $\mathcal{M}$

Now that $\hat\theta$ has been shown to belong to the class $\mathcal{M}$, the next step is to show that it is efficient relative to $\mathcal{M}$. This follows as a consequence of the next result which gives a lower bound to the asymptotic variance of estimators belonging to $\mathcal{M}$.

Theorem 4 : *Let conditions* (A) *and* (B) *be assumed to hold. If T is an estimator belonging to $\mathcal{M}$, then*

$$\int (T^{(1)}[F_{\theta_0}\,;\,x])^2\,dF_{\theta_0}(x)\geqslant\left\{E_{\theta_0}\left(\frac{\partial\log f}{\partial\theta}\right)^2_0\right\}^{-1}. \qquad \dots \quad (4.1)$$

Proof : From the finite Taylor expansion formula and Fisher consistency of T we obtain the relation

$$1=\int T^{(1)}(F_{\theta_0}\,;\,x]\,[\eta f(x,\theta_0)]^{-1}[f(x,\theta_0+\eta)$$

$$-f(x,\theta_0)]dF_{\theta_0}(x)+\eta^{-1}\left(\frac{d^2}{dt^2}\,T[F_{\theta_0}+th]\right)_{t=t(\eta)},$$

where η is sufficiently small, $0\leqslant t(\eta)\leqslant 1$ and $h=F_{\theta_0+\eta}-F_{\theta_0}$. From Condition (B) and Condition (iv) of Definition 1 we have on making $\eta\to 0$

$$1=\int T^{(1)}[F_{\theta_0};x]\left(\frac{\partial\log f(x,\theta)}{\partial\theta}\right)_0\,dF_{\theta_0}(x). \qquad \dots \quad (4.2)$$

The conclusion of the theorem follows immediately on applying Schwarz's inequality to the integral on the right hand side of (4.2).

It seems appropriate at this stage to comment on a definition of efficiency recently introduced by Rao (1962). According to this definition an estimator is said to be efficient if its asymptotic correlation with the derivative of the log likelihood is unity. It follows from (4.2) that a statistic $T\epsilon\mathcal{M}$ is efficient in the new sense if and only if its asymptotic variance is equal to $[n\,i(\theta)]^{-1}$, i.e., if and only if equality holds

in (4.1). Hence the new concept of efficiency and efficiency in the sense of Fisher are equivalent so far as the class $\mathscr{M}$ is concerned.

The results obtained in this paper can be extended to the case when the parameter θ belongs to an open subset of k-dimensional Euclidean space. Finally, we offer a remark in connection with the definition of von Mises functionals of second order. Condition (iii) is very similar to the one imposed in the paper by Filippova (1962) where asymptotic distributions of second and higher order functionals are studied in a different context. Condition (iv) can, if desired, be replaced by a suitable analytical condition on T. For instance, (iv) is satisfied if it is assumed that the second V-derivative, $T^{(2)}[V^{(t)};\ x_1, x_2]$ at $V^t(V^{(t)}\epsilon\tau_T(\theta))$ is bounded uniformly with respect to $V^{(t)}$, x_1 and x_2. However, such a sufficient condition would entail additional restrictions on F_θ if it is to hold for $T = \hat{\theta}$.

REFERENCES

CRAMÉR, H. (1946): *Mathematical Methods of Statistics.* Princeton University Press.

FILIPPOVA, A. A. (1962): A theorem of Mises on the asymptotic behaviour of functionals of empirical distribution functions and its statistical applications (in Russian), *Teoriya Veroyatnostei i ee Primeneniya*, **7**, 26–60.

HUZURBAZAR, V. S. (1948): The likelihood equation, consistency and the maximum of the likelihood function. *Ann. Eugen.* **14**, 185–200.

KALLIANPUR, G. and Rao, C. R. (1955): On Fisher's lower bound to asymptotic variance of a consistent estimate. *Sankhyā*, **15**, 331-342.

RAO, C. R. (1962): Apparent anomalies and irregularities in maximum likelihood estimation. 32nd Session of the *Int. Stat. Inst.* Tokyo, reprinted in *Sankhyā*, Series A, **24**, Part 1, 73-102.

———— (1957): Maximum likelihood estimation for the multinomial distribution. *Sankhyā*, **18**, 139-148.

VON MISES, R. (1947): On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Stat*, **18**, 309–348.

Paper received : March, 1963.

ON THE APPROXIMATION OF DISTRIBUTIONS OF SUMS OF INDEPENDENT SUMMANDS BY INFINITELY DIVISIBLE DISTRIBUTIONS*

By A. N. KOLMOGOROV

Academy of Sciences, U.S.S.R.

Throughout this paper

$$\xi = \xi_1 + \xi_2 + \ldots, + \xi_n$$

is a sum of n independent real random variables,

$$F_k(x) = P(\xi_k < x), \quad H(x) = P(\xi < x)$$

$$G_{\sigma^2}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2\sigma^2}} dx, \sigma > 0$$

$$\epsilon(x) = G_0(x) = \begin{cases} 0 & \text{if } x \leqslant 0 \\ 1 & \text{otherwise,} \end{cases}$$

$v = \{D\}$ is the totality of all infinitely divisible distribution functions $D(x)$ and $c_1, c_2, \ldots$ are positive constants.

The following strengthened versions of two theorems, the weaker forms of which were given in my work (Kolmogorov, 1956) will be proved :

Theorem 1 : *There exists a constant c_1 such that, in the case of identically distributed ξ_k, whatever be $F(x) = F_k(x)$, $k = 1, 2, \ldots, n$, there exists Dev satisfying the inequality*

$$|D(x) - H(x)| \leqslant c_1 n^{-1/3} \qquad \ldots (0.1)$$

for all x.

Theorem 2 : *There exists a c_2 such that, for any $\varepsilon > 0$, $L \geqslant 2l > 0$ the validity of the inequalities*

$$E(x-l) - \varepsilon \leqslant F_k(x) \leqslant E(x+l) + \varepsilon \qquad \ldots (0.2)$$

for all x and $k = 1, 2, \ldots, n$, implies the existence of a Dev for which

$$D(x-L) - \delta \leqslant H(x) \leqslant D(x+L) + \delta \qquad \ldots (0.3)$$

for all x, where

$$\delta = c_2 \max \left(\frac{l}{L} \left(\log \frac{L}{l} \right)^{\frac{1}{2}}, \varepsilon^{1/3} \right). \qquad \ldots (0.4)$$

So goes the history of this problem.

* Translation of a lecture delivered at the Indian Statistical Institute, Calcutta, in April 1962. This paper has been included in *Contributions to Statistics* presented to Professor P. C. Mahalanobis on the occasion of his 70th birthday.

1. From the closure of the class of all infinitely divisible distributions, introduced by Bruno de Finetti (1930), under weak convergence it immediately follows that in the case when the sum

$$\xi_k = \xi_{k1} + \ldots \xi_{kn_k} \qquad \ldots \quad (0.5)$$

$$\lim_{k\to\infty} n_k = \infty$$

of independent summands which are identically distributed within each series, converges weakly the limit distribution is infinitely divisible.

It was tempting to understand this result thus : sum of a large number of identically distributed independent summands has a distribution, approximately normal. However, before my work (Kolmogorov, 1956) such an interpretation was not fully convincing.

Even in the case of a sequence of identically distributed summands

$$\xi_1, \xi_2, \ldots$$

a 'completely different' case was possible according to Doeblin (1959), in which under no normalisation

$$\xi_n = A_n(\xi_1 + \ldots + \xi_n) - B_n$$

and for no sequence $\quad n_1 < n_2 < \ldots < n_k < \ldots$

the distributions of the sums ξ_{n_k} can converge to any distribution other than the degenerate distribution $E(x-a)$. The latter, of course, can be achieved by choosing the multipliers A_n sufficiently small.

Only in 1955 Yu. V. Prohorov proved that, in the case of sequence of independent and identically distributed summands ξ_n there always exists a sequence of infinitely divisible distribution functions

$$D_1(x), D_2(x), \ldots, D_n(x),$$

approximating the distribution $H_n(x)$ of the sums

$$\xi_n = \xi_1 + \ldots, \xi_n$$

in the sense

$$\sup_x |H_n(x) - D_n(x)| \to 0, \qquad \ldots \quad (0.6)$$

as $n\to\infty$. Prohorov's work (1954), however, left open the question whether the convergence in (0.6) is uniform with respect to the choice of the distribution function $F(x)$ of ξ_n.

In terms of the uniform metric

$$\rho(F', F'') = \sup_x |F'(x) - F''(x)|$$

the problem is as follows : do the functions*

$$\psi(n) = \sup_F \rho(H_n, v)$$

converge to zero as $n\to\infty$? My work (Kolmogorov, 1956) gave the answer to this question; it was proved that

$$\psi(n) = O(n^{-1/5}). \qquad \ldots \quad (0.7)$$

*The supremum is taken over all distribution functions F.

In 1960 Prohorov strengthened this result, by showing that

$$\psi(n) = O(n^{-1/3} \log^2 n). \quad \ldots \quad (0.8)$$

Our Theorem 1 states that* $\psi(n) = O(n^{-1/3})$. ... (0.9)

The problem of estimating the function $\psi(n)$ from below naturally arose. Prohorov's students I.P. Tsaragradsky, Prohorov himself and L. D. Meshalkin occupied themselves with such estimates. The latest result of Meshalkin (1961) runs thus :

$$\psi(n) \geqslant c_3 n^{-2/3} (\log n)^{-4}. \quad \ldots \quad (0.10)$$

2. For the case of sums

$$\xi_k = \xi_{k1} + \ldots \ \xi_{kn_k}$$

of random variables, independent within each series, with different distributions, Khintchine (1957), determined sufficient conditions so that the distribution of ξ_k may converge weakly to an infinitely divisible distribution. The conditions are that there exist

$$\varepsilon_k \to 0, \quad l_k \to 0$$

such that the distributions F_{ki} of ξ_{ki} satisfy

$$E(x-l_k) - \varepsilon_k \leqslant F_{ki}(x) \leqslant E(x+l_k) + \varepsilon_k.$$

Our Theorem 2 is an attempt to give an 'uniform' character to this result of Khintchine. The nature of the content of Theorem 2 can be made clearer with the help of "Levy distance,"

$$\rho_h(F', F'') = \inf \varepsilon$$

over all ε satisfying the condition

$$F'(x-\varepsilon n) - \varepsilon \leqslant F''(x) \leqslant F'(x+\varepsilon h) + \varepsilon.$$

It is easy to see that from Theorem 2 we have the following corollary.

Corollary : *If*

$$\sup \ \rho_h(F_{ki}, E) \leqslant \eta$$

then

$$\rho_h(H, v) \leqslant c_4 \eta^{1/3}.$$

3. As is known, the most useful means of proving limit theorems for distributions of sums of a large number of independent variables is the apparatus of characteristic functions. The 'direct' probabilistic method now *in this domain* can only rarely compete with the potentialities of the analytical apparatus of characteristic functions. Our Theorems 1 and 2 are unusual examples of the other situation. The essential element of their proofs is Lemma 1, relating to the 'concentration function' introduced

*After completing this work I came to know that F. M. Kagan, in 1961, discovered the result

$$\psi(n) = O(n^{-1/3} \log n)$$

which is between (0.8) and (0.9). Later this result was reported by F. M. Kagan in the conference on Probability Theory and Mathematical Statistics held in Fergan (September 1962).

by Levy. Theorems of Levy and Doeblin about the properties of concentration function were strengthened by me (cf. Kolmogorov, 1958) specially for proving the earlier versions of Theorems 1 and 2 given in Kolmogorov, (1956). The later development in the estimation of concentration functions belongs to Rogozin (1961a; and 1961b). Rogozin too uses elementary direct probabilistic and set theoretic methods (Theorem 5.5 on subsets of a finite set).

The mathematicians, whose attention to this problem I succeeded in drawing, could not prove theorems of type 1 and 2 without appealing to the just indicated peculiar methods.

Throughout this paper, as in Kolmogorov (1956), the methods of reasoning are essentially those introduced by Doeblin (1959). As is seen from what is said above, the transition from the degree 1/5 to the degree 1/3 in Theorem 1 was done by Prohorov. The removal of the factor $\log^2 n$ in the estimate of Prohorov (0.8) required (a) the use of a more precise estimate of the concentration function obtained by Rogozin and (b) some changes in Prohorov's proof, reflected in the introduction of Lemmas 5 and 6.*

In the proof of Theorem 2 the transition from the degree 1/5 to the degree 1/3 is effected by the techniques borrowed from the work of Prohorov (1960) combined with the use of Lemmas 5 and 6.

4. Besides the distance

$$\rho(F', F'') = \sup_x |F'(x)-F''(x)|$$

it is natural to consider 'the variation distance'

$$\rho_V(F', F'') = \tfrac{1}{2}\operatorname{var}[F'(x)-F''(x)] = \sup_A [F'(A)-F''(A)],$$

where A is an arbitrary Borel measurable set on the real line.

As is known, $$\rho(F', F'') \leqslant \rho_V(F', F'').$$

Hence for the function $$\psi_V(n) = \sup_F \rho_V(F^n, v)$$

we have the inequality $$\psi_V(n) \geqslant \psi(n).$$

However, nothing is known about the asymptotic behaviour of $\psi_V(n)$. It is not even known whether $\psi_V(n) \to 0$ as $n \to \infty$.

1. Eight lemmas

Following Levy, for any distribution function $F(x)$ we introduce the 'concentration function'

$$Q_F(l) = \sup_x [F(x+l+0)-F(x)]$$

Lemma 1 : *If , for* $k = 1, 2, \ldots n, \quad L \geqslant l \geqslant 0$

$$P\{\xi_k \leqslant x_k^{-1}\} = P\{\xi_k \geqslant x_k\} = \tfrac{1}{2}$$

then $$Q_H(L) \leqslant c_5 \frac{L}{l} n^{-\frac{1}{2}}.$$

* The first step was done by F. M. Kagan sometime before my work was written (see footnote on page 161).

Lemma 2 : *If* $\sigma > 0$, $l > 0$, $\eta = c_6 e^{-l^2/2\sigma^2}$, *then for any distribution function* $F(x)$

$$F * G_{\sigma^2}(x-l) - \eta \leqslant F(x) \leqslant F * G_{\sigma^2}(x+l) + \eta.$$

Lemma 3 : *If* $\sigma > 0, \sigma_1 > 0$, *then*

$$|G_{\sigma^2}(x) - G_{\sigma_1^2}(x)| \leqslant c_7 \left| \frac{\sigma^2}{\sigma_1^2} - 1 \right|$$

Lemma 4 : *If* $\int x F(dx) = 0$, $\int x^2 F(dx) = \sigma^2$, $h \geqslant \sigma > 0$

then
$$\sum_{r=-\infty}^{\infty} \sup_{rh \leqslant x \leqslant (r+1)h} |F(x) - E(x)| \leqslant c_8.$$

Lemma 5 : *If* $M\xi_k = 0$, $|\xi_k| \leqslant l$, $D\xi = \sigma^2$, $h \geqslant \sigma > 0$

then
$$\sum_{r=-\infty}^{\infty} \sup_{rh \leqslant x \leqslant (r+1)h} |H(x) - G_{\sigma^2}(x)| \leqslant c_9\, l/\sigma.$$

Lemma 6 : *If* $M\xi_k = 0$, $|\xi_k| \leqslant l$, $D\xi = \sigma^2$, $\sigma_1^2 = \sigma^2 + \sigma_0^2$

then
$$|H * G_{\sigma_0^2}(x) - G_{\sigma_1^2}(x)| \leqslant c_{10}\, l/\sigma_1.$$

Lemma 7 : *For any natural number* n *and* $0 \leqslant p \leqslant 1$

$$\sum_{m=0}^{\infty} \left| c_n^m p^m (1-p)^{n-m} - \frac{(np)^m}{m!} e^{-np} \right| \leqslant c_{11} p.$$

Lemma 8 : *Let**

$$0 \leqslant p_k \leqslant 1$$

$$p_k(m) = \begin{cases} 1-p_k & \text{if } m = 0 \\ p_k & \text{if } m = 1 \\ 0 & \text{if } m > 1 \end{cases}$$

$$q_k(m) = \frac{p_k^m}{m!} e^{-p_k}$$

$$p(\bar{m}) = \prod_k p_k(m_k),\ q(\bar{m}) = \prod_k q_k(m_k).$$

Then
$$\sum_{\bar{m}} |p(\bar{m}) - q(\bar{m})| \leqslant c_{12} \sum_k p_k^2.$$

*Here and everywhere in what follows $\bar{m} = (m_1, \cdots, m_n)$ is an n-dimensional vector and $\sum_{\bar{m}}$ denotes summation over all vectors with non-negative integral components.

Lemmas 2, 3, 5 are proved by simple calculations. Lemma 1 is an immediate corollary of Theorem 1 of Rogozin (1961a). Lemma 7 is proved in the work of Prohorov (1953). Lemma 4 follows from the estimate (ξ has distribution F)

$$|F(x)-E(x)| \leqslant P\{|\xi| \geqslant |x|\} \leqslant \sigma^2/x^2$$

(Chebyshev's inequality).

Lemmas 5 and 6 reduce to the known estimate

$$|H-G_{\sigma_2}| \leqslant c_{13}l/\sigma \qquad \ldots \quad (1.1)$$

which follows from Lyapunov's theorem in the formulation of Esseen (Gnedenko and Kolmogorov 1949, p. 216)

$$|H-G_{\sigma_2}| \leqslant \frac{c_{14}}{\sigma^3}\sum_k M|\xi_k|^3. \qquad \ldots \quad (1.2)$$

Lemma 6 can be obtained from (1.2) if the adjoined normal summand with dispersion σ_0^2 is represented as a sum of a large number of summands with sufficiently small dispersions.

The proof of Lemma 8 is somewhat more complicated.

2. Proof of theorem 1

1. Hereafter we shall assume $n > 1$. It is easy to see that this restriction is immaterial.

2. Further we shall assume that ξ_k are non-decreasing functions

$$\xi_k = F_k^{-1}(\eta_k)$$

of mutually independent variables η_k with uniform distribution on the interval $[0, 1]$.

It is easy to prove that in an extended probability field (Ω, m, P) such quantities η_k exist.

3. Suppose

$$p = n^{-1/3}$$

$$\mu_k = \begin{cases} 0 & \text{if } \ p/2 < \eta_k < 1-p/2 \\ 1 & \text{otherwise,} \end{cases}$$

$$a = M(\xi_k|\mu_k = 0)$$

$$\sigma^2 = D(\xi_k|\mu_k = 0)$$

$$A(x) = P\{\xi_k < x|\mu_k = 0\}; \quad B(x) = P\{\xi_k < x|\mu_k = 1\}.$$

Under the transition from the quantities ξ_k to

$$\xi_k' = \xi_k - a$$

all these constructions remain unchanged. Only instead of a we have $a' = 0$ and the functions $A(x)$ and $B(x)$ become

$$A'(x) = A(x+a); \quad B'(x) = B(x+a).$$

Thus it is enough to consider the case

$$a = 0$$

and to this we shall restrict ourselves hereafter.

4. In the decomposition $F(x) = pB(x)+(1-p)A(x)$ the distribution A is concentrated in the interval $[x^-, x^+]$, where

$$x^- = F^{-1}(p/2), \quad x^+ = F^{-1}(1-p/2).$$

The length of the interval is

$$\lambda = x^+ - x^-.$$

The distribution B is outside the interval $[x^-, x^+]$ and each of the rays $(-\infty, x^-]$ and $[x^+, \infty)$ has probability $\frac{1}{2}$ under B. To the distributions B^m (here and elsewhere power is in the sense of convolution), it is possible to apply Lemma 1, which yields the estimate

$$Q_{B^m}(\lambda) \leqslant c_5\, m^{\frac{1}{2}}. \qquad \ldots \quad (2.1)$$

5. We shall approximate the distribution

$$H = [pB+(1-p)A]^n = \sum_m c_n^m p^m(1-p)^{n-m}B^m * A^{n-m}$$

by an infinitely divisible distribution D in two different cases

$$\text{(A)}\quad \lambda \geqslant \sqrt{n}\,\sigma, \quad ; \quad \text{(B)}\quad \lambda < \sqrt{n}\,\sigma.$$

Case (A): Suppose

$$D = \exp\,[np\,(B-E)] = \sum_m \frac{(np)^m}{m!} e^{-np}B^m,$$

$$H_1 = \sum_m c_n^m\, p^m(1-p)^{n-m}B^m.$$

According to Lemma 7

$$|D-H_1| \leqslant c_{11}p = c_{11} \cdot n^{-1/3} \qquad \ldots \quad (2.2)$$

By Lemma 4 and estimate (2.1) when $h = \lambda$ we have*

$$|B^m * A^{n-m} - B^m| \leqslant \int |A^{n-m}(x-z) - E(x-z)|\, B^m(dz)$$
$$\leqslant Q_{B^m}(\lambda) \sum_r \sup_{r\lambda \leqslant y \leqslant (r+1)\lambda} |A^{n-m}(y)-E(y)| \leqslant c_3c_8\, m^{-\frac{1}{2}}. \qquad \ldots \quad (2.3)$$

Thus

$$|H-H_1| \leqslant \sum c_n^m p^m(1-p)^{n-m}\,|B^m * A^{n-m} - B^m|$$
$$\leqslant c_5c_8\, n^{-1/3} + 2\Sigma', \qquad \ldots \quad (2.4)$$

where

$$\Sigma' = \sum_{m < \frac{1}{2}n^{2/3}} c_n^m p^m(1-p)^{n-m} = P\{\mu < \tfrac{1}{2}n^{2/3}\}.$$

Observing that

$$M\mu = np = n^{2/3} \qquad \ldots \quad (2.5)$$
$$D\mu = np(1-p) \leqslant n^{2/3} \qquad \ldots \quad (2.6)$$

we obtain by Chebyshev's inequality

$$\Sigma' \leqslant P\{|\mu - n^{2/3}| > \tfrac{1}{2}n^{2/3}\} \leqslant 4n^{-2/3}$$

*The dispersion of the distribution A^{n-m} is equal to $(n-m)\sigma^2$ so that the conditions of Lemma 4 are satisfied when $h = \lambda \geqslant \sqrt{n}\,\sigma$.

which, together with (2.4). leads to the inequality

$$|H-H_1| \leqslant (2c_5\, c_8+8)n^{-1/3}. \qquad \dots \quad (2.7)$$

From (2.6) and (2.2) we get (0.1).

Case (B) : Suppose

$$D = \exp[np(B-E)]*G_{n(1-p)\sigma^2}$$

$$= \Sigma \frac{(np)^m}{m!}\, e^{-np} B^m * G_{n(1-p)\sigma^2}$$

$$H_1 = \sum_m c_n^m p^m (1-p)^{n-m} B^m * G_{(n-m)\sigma^2}$$

$$H_2 = \sum_m c_n^m p^m (1-p)^{n-m} B^m * G_{n(1-p)\sigma^2}.$$

By Lemma 7 we have

$$|D-H_2| \leqslant c_{11} n^{-1/3}. \qquad \dots \quad (2.8)$$

The difference $H-H_1$, is estimated with the help of Lemma 5, where we assume now $h = \sqrt{n}\sigma$

$$|B^m * A^{n-m} - B^m * G_{(n-m)\sigma^2}|$$

$$\leqslant \int |A^{n-m}(x-z) G_{(n-m)\sigma^2}(x-z)| B^m(dz)$$

$$\leqslant Q_{B^m}(\sqrt{n}\sigma) \sum_r \sup_{r\sqrt{n}\sigma \leqslant y \leqslant (r+1)\sqrt{n}\sigma} |A^{n-m}(y) - G_{(n-m)\sigma^2}(y)|$$

$$\leqslant c_5 \frac{\sqrt{n}\sigma}{\lambda} m^{-\frac{1}{2}} c_9 \frac{\lambda}{\sqrt{n}\sigma} = c_5 c_9 m^{-\frac{1}{2}}.$$

This estimate is completely analogous to the estimate (2.3) of case (A). Exactly in the same manner as in case (A), we get

$$|H-H_1| \leqslant c_5 c_9 n^{-1/3} + 2\Sigma' \leqslant c_{15} n^{-1/3}. \qquad \dots \quad (2.9)$$

The difference H_1-H_2 is estimated with the help of Lemma 3 :

$$|H_1-H_2| \leqslant c_{16} n^{-1/3} + \Sigma''$$

where

$$\Sigma'' = \sum_{\left|\frac{n-m}{n(1-p)}-1\right| \geqslant \frac{c_{16}}{c_7}} c_n^m p^m (1-p)^{n-m}.$$

It is possible to obtain from (2.5) and (2.6), when $n > 1$ and c_{16} is chosen as above, with the help of Chebyshev's inequality, the estimate

$$\Sigma'' \leqslant c_{17} n^{-1/3}$$

which leads to

$$|H_1-H_2| \leqslant (c_{16}+c_{17}) n^{-1/3}. \qquad \dots \quad (2.10)$$

From (2.8), (2.9), (2.10) we obtain (0.1) which completes the proof of Theorem 1.

3. Proof of theorem 2

1. Without loss of generality, we may consider $\varepsilon < 1$.

2. We shall show that it is enough to consider the case of continuous and strictly increasing functions $F_k(x)$.

Suppose that Theorem 2 has been proved with constant c_2' for the case of continuous and strictly increasing functions. We consider the sum

$$\xi = \xi_1 + \dots \xi_n$$

with arbitrary $F_k(x)$ satisfying (0.2). Let $L > 2l$. We choose l' and L' such that

$$L > L' > 2l' > 2l, \quad \frac{l'}{L'} \geqslant \frac{1}{2}\frac{l}{L} \qquad \dots \quad (3.1)$$

By Lemma 2 it is possible to choose such a small σ_0 that, for any distribution function $F(x)$ the inequalities

$$F*G_{\sigma_0^2}(x-\lambda)-\varepsilon \leqslant F(x) \leqslant F* \leqslant G_{\sigma_0^2}(x+\lambda)+\varepsilon \qquad \dots \quad (3.2)$$

$$F*G_{n\sigma_0^2}(x-\Lambda)-\delta' \leqslant F(x) \leqslant F*G_{n\sigma_0^2}(x+\Lambda)+\delta' \qquad .. \quad (3.3)$$

where

$$\lambda = l-l', \quad \Lambda = L-L'$$

$$\delta' = c_2' \max\left[\frac{l'}{L'}\left(\log\frac{L'}{l'}\right)^{\frac{1}{2}}, (2\varepsilon^{1/3})\right]$$

are fulfilled. Let

$$F_k' = F_k*G_{\sigma^2},$$

By (0.2) and (3.2) we get

$$E(x-l')-2\varepsilon \leqslant F_k'(x) \leqslant E(x+l')+2\varepsilon.$$

Since the function F_k' are continuous and strictly increasing, there exists an infinitely divisible distribution D', for which

$$D'(x-L')\delta' \leqslant H'(x) \leqslant D'(x+L')+\delta'. \qquad \dots \quad (3.4)$$

Observing that

$$H' = H*G_{n\sigma_0^2}$$

from (3.3) applied to $F(x) = H(x)$ and (3.4), we obtain

$$D'(x-L)-2\delta' \leqslant H(x) \leqslant D'(x+L)+2\delta' \qquad \dots \quad (3.4a)$$

since, by (3.1)

$$2\delta' \leqslant 2c_2' \max\left[\frac{l'}{L'}\left(\log\frac{L'}{l'}\right)^{\frac{1}{2}}, (2\varepsilon)^{1/3}\right]$$

$$\leqslant c_2 \max\left[\frac{l}{L}\left(\log\frac{L}{l}\right)^{\frac{1}{2}}, \ \varepsilon^{1/3}\right].$$

when c_2 is chosen as above, (0.3) follows from (3.4a.)

3. In accordance with 2 we shall hereafter consider the case when $F_k(x)$ are continuous and strictly increasing. Then the function

$$\lambda_k(p) = F_k^{-1}(1-p/2) - F_k^{-1}(p/2)$$

will be well-defined for all p, $0 < p < 1$, continuous and strictly decreasing. It takes all positive values. Therefore the inverse function $p_k(\lambda)$ is defined on $0 < \lambda < \infty$ as a continuous strictly decreasing function taking all values in the interval $1 > p > 0$.

The function

$$s(\lambda) = \sum_k p_k(\lambda)$$

is also continuous and strictly decreasing. It takes all values $n > s > 0$. Thus, when $0 < \varepsilon < 1$, there exists a unique solution λ_0 for the equation $s(\lambda) = \varepsilon^{-2/3}$.

1. Let

$$\Lambda = \begin{cases} \lambda_0 & \text{if } \lambda_0 \geqslant l \\ l & \text{if } \lambda_0 < l \end{cases}$$

$$p_k = p_k(\Lambda),\ s = s(\Lambda) = \sum_k p_k$$

$$x_k^- = F_k^{-1}(p_{k/2}),\ x_k^+ = F_k^{-1}(1-p_{k/2})$$

$$\mu_k = \begin{cases} 0 & \text{if } x_k^- < \epsilon_k < x_k^+ \\ 1 & \text{otherwise} \end{cases}$$

$$a_k = M(\xi_k/\mu_k = 0),$$
$$\sigma_k^2 = D(\xi_k/\mu_k = 0),$$
$$\sigma^2 = \sum_k (1-p_k)\,\sigma_k^2.$$

Putting

$$A_k(x) = P\{\xi_k < x \mid \mu_k = 0\} : \quad B_k(x) = P\{\xi_k < x \mid \mu_k = 1\},$$

we represent $F_k(x)$ in the form

$$F_k(x) = p_k\, B_k(x) + (1-p_k)\, A_k(x)$$

where the distribution A_k is concentrated in the interval $[x_k^-, x_k^+]$ and B_k outside this interval such that the rays $(-\infty\ x_k^-]$ and $[x_k^+, \infty)$ have probability $\frac{1}{2}$ each.

Using the notations of Lemma 8 and putting*

$$B(\bar{m}) = \prod_k B_k^{m_k},\ A(\bar{m}) = \prod_k A_k^{1-m_k}$$

we obtain $$H = \prod_h [p_k B_k + (1-p_k) A_k] = \sum_{\bar{m}} p(\bar{m}) B(\bar{m}) * A(\bar{m}).$$

* $B(\bar{m})$ is defined for arbitrary non-negative m_k but $A(\bar{m})$ only in the case when m_k's take the value 0 or 1.

The construction of the approximating infinitely divisible distribution will be different in three cases (A), (B) and (C) :

A	B	C
$\lambda_0 \geqslant l$ $\lambda_0 \geqslant \sigma$	$\lambda_0 \geqslant l$ $\lambda_0 < \sigma$	$\lambda_0 < l$
$\Lambda = \lambda_0$ $s = \varepsilon^{-2/3}$	$\Lambda = \lambda_0$ $s = \varepsilon^{-2/3}$	$\Lambda = l$ $s = \varepsilon^{-2/3}$

5. Since $\Lambda \geqslant l$ always

$$p_k = p_k(\Lambda) \leqslant p_k(l) \leqslant \varepsilon. \qquad \dots \quad (3.5)$$

It is only here that we appeal to conditions (0.2) of the theorem in our proof. Since the definition of $p_k(\Lambda)$ and all other essential constructions are invariant with respect to shifts

$$\xi_k' = \xi_k - c_k$$

we can restrict ourselves to the case

$$a_k = 0.$$

6. We shall make some more simplifications which we need in future.

Let

$$\tau = \sum_k \mu_k, \qquad t(\overline{m}) = \sum_k m_k$$

τ is equal to the number of times ξ_k lies in the interval $x_k^- < \xi_k < x_k^+$. It is easy to verify that

$$M\tau = s$$

$$D\tau = \sum_k p_k(1-p_k) < s.$$

Thus, by Chebyshev's inequalities

$$P\{|\tau - s| \geqslant c\} = \sum_{|t(\overline{m})-s| \geqslant c} p(\overline{m}) < \frac{s}{c^2}. \qquad \dots \quad (3.6)$$

7. Let further

$$\xi = \sum_k (1-\mu_k)\xi_k.$$

It is the sum of all those ξ_k which lie in the interval $x_k^- < \xi_k < x_k^+$. Because of the assumption $a_k = 0$ for all $\overline{m}$

$$M\xi = 0 : \qquad M(\xi \,|\, \bar{\mu} = \overline{m}) = 0.$$

Further we are interested in the conditional dispersions

$$\sigma^2(\overline{m}) = D(\xi / \bar{\mu} = \overline{m}).$$

$$\rho^2 = \sigma^2(\mu)$$

For the random variable

it is easy to verify that

$$M\rho^2 = \sigma^2.$$

Since

$$\Sigma(1-p_k)\sigma_k^2 = \sigma^2$$

$$\sigma_k^2 \leqslant \frac{\Lambda^2}{4}$$

$$p_k = p_k(\Lambda) \leqslant p_k(l) \leqslant \varepsilon$$

we have

$$D\rho^2 \leqslant \frac{1}{4}\,\sigma^2\Lambda^2\varepsilon.$$

Thus

$$P(|\rho^2-\sigma^2| \geqslant c) = \sum_{|\sigma^2(m)-\sigma^2| \leqslant c} p(\overline{m}) \leqslant \frac{\sigma^2\Lambda^2\varepsilon}{4c^2}. \quad \ldots \quad (3.7)$$

8. Finally, we observe that by Lemma 1

$$Q_{B(\overline{m})}(L) \leqslant c_5 \frac{L}{\lambda_0}\, t(\overline{m}). \quad \ldots \quad (3.8)$$

We immediately pass on to the proof of Theorem 2 in cases (A), (B) and (C).

Case (A) : In this case

$$H = \prod_k [p_kB_k+(1-p_k)A_k] = \sum_{\overline{m}} q(\overline{m})\, B(\overline{m}) * A(\overline{m})$$

is approximated by

$$D = \exp \sum_k p_k(B_k-E) = \sum_{\overline{m}} q(\overline{m})\, B(\overline{m}).$$

In order to pass on to D from H we further consider

$$H_1 = \sum_{\overline{m}} p(\overline{m})\; B(\overline{m}).$$

By Lemma 8 and (3.5)

$$|D-H_1| \leqslant \sum_{\overline{m}} |p(\overline{m})-q(\overline{m})| \leqslant c_{12} \sum_k p_k^2$$

$$\leqslant c_{12}\varepsilon \sum_k p_k \leqslant c_{12}\, \varepsilon^{1/3}. \quad \ldots \quad (3.9)$$

On the other hand, by Lemma 4 when $h = \Lambda = \lambda_0 \geqslant \sigma$, and (3.8)

$$|B(\overline{m})*A(\overline{m})-B(\overline{m})| \leqslant \int |A(\overline{m})(x-z)-E(x-z)|\, B(\overline{m})(dz)$$

$$\leqslant Q_{B(\overline{m})}(\lambda_0) \sum_r \sup_{r\lambda_0 \leqslant y \leqslant (r+1)\lambda_0} |A(\overline{m})(y)-E(y)|$$

$$\leqslant c_5c_8[t(\overline{m})]^{-\frac{1}{2}}. \quad \ldots \quad (3.10)$$

Thus

$$|H-H_1| \leqslant \sum_{\overline{m}} p(\overline{m})\, |B(\overline{m})*A(\overline{m})-B(\overline{m})|$$

$$\leqslant \sqrt{2}\; c_5\, c_8\varepsilon^{-1/3}+2\Sigma' \quad \ldots \quad (3.11)$$

where

$$\Sigma' = \sum_{t(\overline{m}) < \frac{1}{2}\varepsilon^{2/3}} p(m).$$

Observing that in our case $s = \varepsilon^{2/3}$, we have by (3.6)

$$\Sigma' < 8\varepsilon^{-2/3} < 4\varepsilon^{1/3}$$

i.e., from (3.11) it follows that

$$|H-H_1| \leqslant (\sqrt{2}c_5c_8+8)\varepsilon^{1/3}. \qquad \ldots \quad (3.12)$$

From (3.9) and (3.12) we obtain

$$|H-D| \leqslant c_{18}\varepsilon^{1/3}. \qquad \ldots \quad (3.13)$$

Case (B) : In this case we put

$$D = \exp \Sigma \; p_k(B_k-E)^*G_{\sigma^2} = \sum_{\overline{m}} q(\overline{m}) \, B(\overline{m})^*G_{\sigma^2},$$

$$H_1 = \sum_{\overline{m}} p(\overline{m}) \, B(\overline{m})^*G_{\sigma^2} \, (\overline{m}),$$

$$H_2 = \sum_{\overline{m}} p(\overline{m}) \; B(\overline{m})^*G_{\sigma^2}.$$

The inequality
$$|D-H_2| \leqslant c_{12} \, \varepsilon^{1/3} \qquad \ldots \quad (3.14)$$

is obtained exactly in the same way as (3.9) was in case (A).

By Lemma 5 when $h = \sigma > \lambda = \lambda_0$ and (3.8) we obtain

$$\begin{aligned}
&|B(\overline{m})^*A(\overline{m})-B(\overline{m})^*G_{\sigma^2}(\overline{m})| \\
&\qquad \leqslant \int |A(\overline{m})(x-z)-G_{\sigma^2(\overline{m})} \; (x-z)| \, B(\overline{m})(dz) \\
&\qquad \leqslant Q_{B(\overline{m})} \, (\sigma) \; \sum_r \sup_{r\sigma \leqslant y \leqslant (r+1)\sigma} |A(\overline{m})-G_{\sigma^2(\overline{m})}| \qquad \ldots \quad (3.15) \\
&\qquad \leqslant c_5 \, \sigma/\lambda_0[t(\overline{m})]^{-1/2} \, . \, c_9 \; \lambda_{0/\sigma} = c_5c_9[t(\overline{m})]^{-\frac{1}{2}}.
\end{aligned}$$

Exactly in the same way as (3.12) was obtained from (3.10) in case (A), we obtain from (3.15)
$$|H-H_1| \leqslant (c_5c_9+8)\varepsilon^{1/3}. \qquad \ldots \quad (3.16)$$

It remains to estimate the difference H_1-H_2. By Lemma 3

$$|G_{\sigma_2}-G_{\sigma_2(\overline{m})}| \leqslant c_{19}\varepsilon^{\frac{1}{3}}$$

if

$$\left| \frac{\sigma^2(\overline{m})}{\sigma^2} - 1 \right| \leqslant \frac{c_{19}}{c_7} \, .$$

With the help of (3.7) and taking into account that now $\Lambda < \sigma$ we obtain the estimate

$$\Sigma'' = \Sigma \; p(\overline{m}) \leqslant c_{20}\varepsilon^{1/3}$$

$$\left| \frac{\sigma^2(\overline{m})}{\sigma^2} - 1 \right| \leqslant \frac{c_{19}}{c_7} \, . \qquad \ldots \quad (3.17)$$

Thus, analogous to the deduction of (2.10) in the proof of Theorem 1, we obtain

$$|H_1-H_2| \leqslant (c_{19}+c_{20})\varepsilon^{1/3}.$$

From (3.14), (3.16) and (3.17) it follows that

$$|H-D| \leqslant c_{21}\,\varepsilon^{1/3}. \quad \ldots \quad (3.18)$$

The inequalities (3.12) and (3.18) show that in cases (A) and (B) the estimate (0.3), constituting the content of Theorem 2 can be changed to the stronger one :

$$D(x)-c_{22}\varepsilon^{1/3} \leqslant H(x) \leqslant D(x)+c_{22}\varepsilon^{1/3}. \quad \ldots \quad (3.19)$$

Case (C) : In this case, getting an estimate of the type (3.19) does not work. We suppose that

$$\sigma_0 = \frac{1}{\sqrt{2}}\,L\left(\log\frac{L}{l}\right)^{-\frac{1}{2}} \quad \ldots \quad (3.20)$$

and introduce an auxiliary distribution

$$H' = H^*G_{\sigma_0^2}.$$

By Lemma 2, if

$$\eta = c_6 e^{-L^2/2\sigma_0^2} = l/L, \quad \ldots \quad (3.21)$$

then

$$H'(x-L)-\eta \leqslant H(x) \leqslant H'(x+L)+\eta. \quad \ldots \quad (3.22)$$

Further we shall show that the infinitely divisible law

$$D = \sum_m \bar{q}(m)\,B(m)^*G_{\sigma_1^2},\quad \sigma_1^2 = \sigma^2+\sigma_0^2$$

satisfies the inequality

$$|D-H'| \leqslant c_{23}\left[\,\varepsilon^{1/3}+\frac{l}{L}\left(\log\frac{L}{l}\right)^{\frac{1}{2}}\right]. \quad \ldots \quad (3.23)$$

Together with (3.21) and (3.22), (3.23) yields (0.3).

It remains to prove (3.23). To this end we introduce the distributions

$$H_1' = \sum_m p\,(\bar{m})\;B(\bar{m})^*G_{\sigma_1^2(\bar{m})};\quad \sigma_1^2(\bar{m}) = \sigma^2(\bar{m})+\sigma_0^2$$

$$H_2' = \sum_m\, p(\bar{m})\;B(m)^*G_{\sigma_1^2}.$$

The proof of the inequality

$$|D-H_2'| \leqslant c_{12}\varepsilon^{1/3} \quad \ldots \quad (3.24)$$

is exactly the same as the proof of (3.9) and (3.14) in cases (A) and (B). But now

$$s = \sum_k p_k \varepsilon^{-2/3}$$

which leads to the estimate

$$\sum_k p_k^2 \leqslant \varepsilon \sum_k p_k < \varepsilon^{1/3}$$

By Lemma 3

$$|H_1' - H_2'| \leqslant c_{24}\varepsilon^{1/3} + \Sigma''' p(\bar{m}) \qquad \ldots \quad (3.25)$$

where Σ''' is taken over those $p(\bar{m})$ for which

$$\left| \frac{\sigma_1^2(\bar{m})}{\sigma_1^2} - 1 \right| > \frac{c_{24}}{c_7}. \qquad \ldots \quad (3.26)$$

With the above choice of c_{24}, it follows from (3.25) that

$$|\sigma_1^2(\bar{m}) - \sigma_1^2| = |\sigma^2(\bar{m}) - \sigma^2)| > \sigma_1^2 \varepsilon^{1/3}.$$

Thus by (3.7) and the inequalities the second inequality is obtained from (3.20).

$$\sigma^2 < \sigma_1^2\, \Lambda = l \leqslant \tfrac{1}{2} L \leqslant \sigma_0$$

we have

$$\Sigma''' \leqslant \frac{\sigma^2 \Lambda^2 \varepsilon}{4\varepsilon^{2/3}\sigma_1^4} \leqslant \frac{1}{4}\, \varepsilon^{1/3}.$$

Thus from (3.25) we obtain

$$|H_1' - H_2'| \leqslant (c_{24}+1)\varepsilon^{1/3}. \qquad \ldots \quad (3.27)$$

Finally, by Lemma 6 and (3.20) we obtain

$$|A(\bar{m})_* G_{\sigma_0^2} - G_{\sigma_1^2(\bar{m})}| \leqslant c_{10}\, l/\sigma_0 \leqslant \sqrt{2}\, c_{10} \left(\frac{l}{L} \left(\log \frac{L}{l} \right)^{\frac{1}{2}} \right),$$

whence

$$|H' - H_1'| \leqslant \sqrt{2}\, c_{10}\, \frac{l}{2} \left(\log \frac{L}{l} \right)^{\frac{1}{2}} \qquad \ldots \quad (3.28)$$

From (3.24), (3.27) and (3.28), (3.23) follows immediately. This completes the proof of Theorem 2.

References

DOEBLIN, WOLFGANG, (1959): Sur les sommes d'un grand nombre de variables aléatoires indépendantes. *Bull. Sci. Math.* **63**, 23–32, 35–64.

FINETTI, BRUNO DE (1930): Le funzioni caracteristische di legge istantnea. *Rendiconti dei Lincei*, **12**, 278–282.

GNEDENKO, B. V. and KOLMOGROV, A. N. (1949): *Limit Distributions for Sums of Independent Random Variables*, Gostekhizdat, Moscow-Leningrad.

KHINTCHINE, A. YA. (1957): Zur Theorie der un beschrankt Teilbaren Verteilungsgesetze. *Mat. Sbornik*, **2**, 79–119.

KOLMOGOROV, A. N. (1956) : Two uniform limit theorems for sums of independent components. *Theory of Probability and its Applications*, **1**, 427–430.

——— (1958) : Sur les proprietés des functions de concentration de M.P. Lévy. *Annals de l'Institut Henri Poincaré*, **16**, fak 1, 27–34.

MESHALKIN, L. D. (1961) : On the approximation of distributions of sums by infinitely divisible laws. *Theory of Probability and its Applications*, **6**, 257–275.

PROHOROV, YU. V. (1953) : Asymptotic behaviour of the binomial distribution. *Uspekhi Math. Nauk*, **8**, N. S. (55), 135–142.

——— (1954) : On sums of identically distributed random variables *DAN USSR*, **105**, 645–647.

——— (1960) : On the uniform limit theorems of A. N. Kolmogorov. *Theory of Probability and its Applications*, **5**, 103–113.

ROGOZIN, B. A. (1961a) : On an estimate of the concentration function. *Theory of Probability and its Applications*, **6**, 103–105.

——— (1961b) : On the increase of dispersion of the sum of independent random quantities. *Theory of Probability and its Applications*, **6**, 106–108.

APPLICATIONS OF CHARACTERISTIC FUNCTIONS IN STATISTICS*

By EUGENE LUKACS

The Catholic University of America

SUMMARY. Applications of characteristic functions, such as the distribution problem of statistics and various characterization and regression problems are discussed. In section 6 the binomial and the negative binomial population are characterized by a regression property.

1. Introduction

Characteristic functions and generating functions were originally developed for the solution of certain problems in probability theory, in particular for the study of the limit distributions of sums of independent random variables. In the present paper we wish to show that there exists also a large variety of problems in mathematical statistics which can be approached by the method of characteristic functions. A certain familiarity with the theory of characteristic functions and their applications is therefore a useful asset for the mathematical statistician.

We do not intend to give here a comprehensive survey of the applications of characteristic functions in mathematical statistics. Our aim is more modest, we wish to indicate only a few typical results in this area. The present paper is largely expository, but we will occasionally mention open problems and will also discuss, in Section 6, a characterization of two discrete distributions which is probably new.

In numerous branches of science one is interested in quantities which are subject to random fluctuations. Series of independent measurements are performed under identical conditions in order to obtain data which should permit conclusions concerning the phenomenon under investigation. The result of the measurements can be treated as independently and identically distributed random variables, say $X_1, X_2, \ldots, X_n$. In such a situation we say that the observations $X_1, X_2, \ldots, X_n$ form a sample of size n drawn from a certain population. We will call the common distribution function of these random variables the population distribution function. The statistician introduces functions of the observations and studies their properties in order to arrive at conclusions concerning the population. These functions cannot be quite arbitrary, they are only useful if they are also random variables. This consideration necessitates the restriction to single valued and measurable functions of the observations. These functions are called statistics. The statistical problems which we intend to discuss in this paper will usually be formulated in terms of certain statistics.

*This paper was prepared with support from the National Science Foundation grant NSF-GP-96.

A few remarks concerning our notation follow :

We will use capital letters taken from the end of the alphabet for random variables; distribution functions will be denoted by $F(x)$, $G(x)$, etc., while $f(t)$, $g(t)$, etc. are the notations for the corresponding characteristic functions. Random variables, distribution functions and characteristic functions may be written with indices which will indicate their connection.

2. THE DISTRIBUTION PROBLEM OF STATISTICS

Let $X_1, X_2, \ldots, X_n$ be a sample of size n drawn from a population with population distribution function $F(x)$ and denote a statistic by $S = S(X_1, X_2, \ldots, X_n)$. The statistic S is a random variable; we write $F_S(x)$ for its distribution function and

$$f_S(t) = \mathcal{E}(e^{itS}) = \int_{-\infty}^{\infty} e^{ixt} dF_S(x)$$

for its characteristic function. The problem of determining the distribution function of a statistic for a given population distribution function is called the distribution problem of statistics.

An important result from the theory of characteristic functions permits to determine the characteristic function of a statistic S by means of the relation[1]

$$f_S(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\,[itS(x_1, x_2, \ldots, x_n)] dF(x_1) dF(x_2) \ldots dF(x_n). \qquad \ldots \quad (2.1)$$

The corresponding distribution function $F_S(x)$ is obtained by means of the inversion formula

$$F_S(x+h) - F_S(x) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{1-e^{ith}}{it}\, e^{-itx} f_S(t) dt \qquad \ldots \quad (2.2)$$

provided that $[x, x+h]$ is a continuity interval of $F_S(x)$. If $f_S(t)$ is absolutely integrable over $(-\infty, +\infty)$ then

$$F'_S(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f_S(t) dt \qquad \ldots \quad (2.2a)$$

yields the frequency function of S. It follows from (2.1) and (2.2) that the distribution $F_S(x)$ of the statistic S is uniquely determined by the population distribution function $F(x)$. Thus we have obtained in principle a solution of the distribution problem of statistics.[2] However, the practical usefulness of this method depends on the possibility to evaluate the integrals (2.1) and (2.2) [respectively (2.2a)].

We mention next a few examples in which the distribution problem can be treated in this manner.

[1] See Laha and Lukacs (1963), Theorem 1.5.8.

[2] The method of characteristic functions was systematically applied for the solution of distribution problems by Kullback (1934) and (1936).

Let $X_1, X_2, \ldots, X_n$ be a sample from a certain population with characteristic function $f(t)$ and write $\overline{X} = \sum_{j=1}^{n} (X_j/n)$ for the sample mean. The well-known convolution theorem permits the derivation of an explicit formula for the characteristic function $f_{\overline{X}}(t)$ of $\overline{X}$. One obtains

$$f_{\overline{X}}(t) = \left[f\left(\frac{t}{n}\right)\right]^n. \quad \ldots \quad (2.3)$$

The distribution function $F_{\overline{X}}(x)$ is uniquely determined by (2.3), the inversion formula yields

$$F_{\overline{X}}(x+h) - F_{\overline{X}}(x) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{1-e^{ith}}{it} e^{-itx}\left[f\left(\frac{t}{n}\right)\right]^n dt \quad \ldots \quad (2.4)$$

provided that $[x, x+h]$ is a continuity interval of $F_{\overline{X}}(x)$. This formula is of practical importance whenever it is possible to evaluate the integral in (2.4). If the population distribution is absolutely continuous, then this is also true for $F_{\overline{X}}(x)$; if in addition $f(t)$ is absolutely integrable then one can determine the probability density of $\bar{X}$ by means of the Fourier inversion formula (2.2a) and obtains

$$F'_{\overline{X}}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \left[f\left(\frac{t}{n}\right)\right]^n dt. \quad \ldots \quad (2.4a)$$

We use a different inversion* formula in case the characteristic function $f(t)$ of an absolutely continuous population distribution function is not absolutely integrable, we obtain then

$$F'_{\overline{X}}(x) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} e^{-itx} \left(1-\frac{|t|}{T}\right) f(t)\, dt. \quad \ldots \quad (2.4b)$$

We obtain already from formula (2.3)—even without using an inversion formula—interesting results concerning the distribution of the sample mean in several populations. We see immediately that the sample mean of a sample drawn from a normal population $N(\mu, \sigma^2)$ has also a normal distribution with mean μ and variance σ^2/n. Similarly, in a Gamma population with parameters θ and λ, the sample mean has also a Gamma distribution with parameters $n\theta$ and $n\lambda$. On the other hand one can see that the Cauchy population has the interesting property that the distribution function of the sample mean is identical with the population distribution function. Other populations for which the distribution of the sample mean can be studied in this manner are the rectangular and the Bessel function populations [see Laha and Lukacs (1963)]. The frequency function of the sample mean in Pearson type I, type VII and type II populations was studied by Irwin (1927, 1929, 1930). The distribution of the geometric mean of a sample can be investigated in a similar manner. The geometric mean of a sample drawn from a rectangular population was studied by Schulz-Arenstorff and Morelock (1959). The distribution problem for the geometric mean of a Gamma population was treated by Kullback (1934).

* See for instance Goldberg (1961), Theorem 6C. We note that (2.4b) is valid also when $f(t)$ is not absolutely integrable.

Let $X_1, X_2, \ldots, X_n$ be a sample from a certain population and let $Q = Q(X_1, X_2, \ldots, X_n)$ be a statistic. It is then interesting to find conditions which assure that the distribution function of Q determines the population distribution function. For instance, if Q is a linear statistic which is normally distributed, then the population distribution function is necessarily also normal. Linnik (1956) considered a rather general statistic Q (subject to certain restrictions) and assumed also that the population distribution function satisfies certain conditions. He derived several theorems which deal with situations in which the knowledge of the distribution function of Q assures that thə population distribution function belongs to a certain family of distributions.

The method of characteristic functions is also very useful in the study of distribution problems of normal populations. The distributions of many important statistics can be derived by using (2.1) and (2.2). We mention here only as examples the chi-square and the non-central chi-square distributions. Characteristic functions can also be used to derive necessary and sufficient conditions which assure that a quadratic polynomial in independently distributed normal variables have a chi-square (respectively a non-central chi-square) distribution. It is also possible to obtain by the method of characteristic functions the distribution of the quotient X/Y of two independent random variables X and Y whose distributions are known. Considering the quotient of two normally distributed random variables we obtain the Cauchy distribution. The distribution of the quotient of two random variables, each of which has a Gamma distribution can be studied in a similar manner; the F-distribution and Student's distribution are obtained as particular cases. The converse problem of determining the families of random variables whose quotient follows these laws was studied by several authors. Mauldon (1956), Steck (1958), Kotlarski (1960) gave examples of random variables whose quotient follows the Cauchy law. Laha (1959a) and (1959b) obtained some general theorems in this area and started also a study* of similar problems related to the F-distribution.

Distribution problems in multivariate populations can also be treated by the method of characteristic functions. We mention here only one example :

Theorem 2.1 : *Let $\boldsymbol{X} = (X_1, \ldots, X_p)$ be a random vector with non-singular p-variate normal distribution with mean vector $(0, 0, \ldots, 0)$ and variance-covariance matrix Λ and let $\boldsymbol{M}$ be a symmetric $p \times p$ matrix and $Q = \boldsymbol{XMX'}$ be a quadratic statistic. The characteristic function of Q is then*

$$f_Q(t) = \mathcal{E}\,(e^{itQ}) = \mathcal{E}\,(e^{it\boldsymbol{XMX'}}) = |\boldsymbol{I} - 2it\Lambda \boldsymbol{Q}|^{\frac{1}{2}}$$

where $\boldsymbol{I}$ is a $p \times p$ matrix whose diagonal elements are 1 *while all other elements are zero.*

It follows then that $\boldsymbol{XMX'}$ has a chi-square distribution with r degrees of freedom if, and only if, $\Lambda \boldsymbol{M}$ has rank r and has exactly r eigenvalues equal to 1.

* Not yet published.

3. Characterization problems—linear statistics

We consider n independently, but not necessarily identically, distributed random variables $X_1, X_2, \ldots, X_n$ with distribution functions $F_1(x), F_2(x), \ldots, F_n(x)$. If the random variables $X_1, X_2, \ldots, X_n$ are identically distributed with common distribution function $F(x)$, we can consider $X_1, X_2, \ldots, X_n$ as a sample of size n, taken from a population with population distribution function $F(x)$. We introduce measurable functions of the random variables and call them statistics (even if the $X_1, X_2, \ldots, X_n$ are not identically distributed). In this section we discuss the problem of determining the distribution functions $F_1(x), F_2(x), \ldots, F_n(x)$ [respectively $F(x)$] from suitable assumptions concerning the distribution of two linear statistics, $L_1 = a_1X_1+a_2X_2+\cdots+a_nX_n$ and $L_2 = b_1X_1+b_2X_2+\cdots+b_nX_n$. The following two assumptions concerning L_1 and L_2 were considered extensively :

(a) L_1 and L_2 are independently distributed,

(b) L_1 and L_2 are identically distributed.

(a) The problem of two independently distributed linear statistics has a long history; it was solved in full generality by Darmois (1953) and Skitovich (1953, and 1954). These authors formulated the problem in terms of characteristic functions and used the method of finite differences, a technique which was suggested independently by Darmois (1951) and Gnedenko (1948). We state here only the result and refer the reader for a proof either to the papers by Darmois (1953) and Skitovich (1954) mentioned above or to Laha and Lukacs (1963).

Theorem 3.1 (Skitovich) : *Let $X_1, X_2, \ldots, X_n$ be n independently but not necessarily identically distributed random variables. Suppose that the two linear forms $L_1 = a_1X_1+a_2X_2+\cdots+a_nX_n$ and $L_2 = b_1X_1+b_2X_2+\cdots+b_nX_n$ are independently distributed. Then each random variable which has non-zero coefficient in both forms is normally distributed.*

(b) The problem of identically distributed linear statistics (even in infinitely many random variables) was first studied in an important paper by Marcinkiewicz (1939). A number of very interesting general results were obtained by Linnik (1953). We state here only one of his principal results. To formulate it we introduce the following terminology. Let $L_1 = a_1X_1+a_2X_2+\cdots+a_nX_n$ and $L_2 = b_1X_1+b_2X_2+\cdots+b_nX_n$, we introduce the entire function $G(z)$ of the complex variable z by writing

$$G(z) = |a_1|^z+|a_2|^z+\cdots+|a_n|^z-|b_1|^z-|b_2|^z-\cdots-|b_n|^z$$

and call it the determining function. We are now in a position to formulate Linnik's result :

Theorem 3.2 (Linnik) : *Let $X_1, X_2, \ldots, X_n$ be n independently and identically distributed random variables with common distribution function $F(x)$. Suppose that $L_1 = a_1X_1+a_2X_2+\cdots+a_nX_n$ and $L_2 = b_1X_1+b_2X_2+\cdots+b_nX_n$ are two linear statistics such that*

$$\max_{1\leqslant j\leqslant n} |a_j| \neq \max_{1\leqslant j\leqslant n} |b_j| .$$

For the equivalence of the two statements

(I) $F(x)$ *is a normal distribution*

(II) L_1 *and* L_2 *are identically distributed*

it is necessary and sufficient that the following five conditions be satisfied,

(i) $a_1+a_2+\cdots+a_n = b_1+b_2+\cdots+b_n$,

(ii) $G(2) = 0$,

(iii) *all zeros of* $G(z)$ *which are integers and are divisible by* 4 *are simple roots,*

(iv) *all positive roots of* $G(z)$ *which are even integers of the form* $4n+2$ *have a multiplicity not exceeding* 2. *If there exists such a double root, then it is unique and is the greatest positive root of* $G(z)$,

(v) *the determining function* $G(z)$ *can have at most one odd integer positive real root* γ. *If such a root exists then it is simple and* $\left[\frac{\gamma}{2}\right]$ *is odd.*

Theorems 3.1 and 3.2 yield characterizations of the normal distribution.

4. Characterization problems—polynomial statistics

Let $X_1, X_2, \ldots, X_n$ be a sample from a normal population and write $\overline{X} = \sum_{j=1}^{n}(X_j/n)$ for the sample mean and $s^2 = \sum_{j=1}^{n}(X_j-\overline{X})^2/n$ for the sample variance. Using the methods discussed in Section 2, one can derive the joint distribution of $\overline{X}$ and s^2 and see that $\bar{X}$ and s^2 are independently distributed. This fact is of great importance in the derivation of Student's test and was first noted by Fisher (1925). The converse of R. A. Fisher's result is also true and led to the first characterization of a population by the independence of two statistics which we formulate next.

Theorem 4.1 : *Let* $X_1, X_2, \ldots, X_n$ *be a sample from a certain population and denote the sample mean by* $\overline{X}$ *and the sample variance by* s^2. *The statistics* $\overline{X}$ *and* s^2 *are independently distributed if, and only if, the population is normal.*

Theorem 4.1 was first proven under certain unnecessary restrictions which were gradually removed by several authors. For its history we refer the reader to Laha and Lukacs (1963) or Lukacs (1956).

Theorem 4.1 was generalized in two directions : Firstly (Laha, 1956) by considering the independence of the mean and a homogeneous quadratic statistic; secondly by considering polynomials of degree exceeding two. Let k_p be Fisher's k-statistics* of order p; it can be shown that the normal populationi s characterized by the independence of k_p and $\overline{X}$.

We mention next a similar result. We denote the sample central moment of order p by $m_p = \sum_{j=1}^{n}(X_j-\overline{X})^p/n$ and suppose that $(p-1)!$ is not divisible by $(n-1)$. The population is normal if, and only if, m_p and $\overline{X}$ are independently distributed.

* The k-statistic k_p of order p ($p \geqslant 1$, integer) is a symmetric and homogeneous polynomial statistic of order p such that $\mathcal{E}(k_p) = \kappa_p$ where κ_p is the p-th cumulant of the population distribution function. It is easily seen that $k_1 = \overline{X}$ and $k_2 = n\, s^2/(n-1)$. The k-statistics of order greater than 2 are not proportional to central moments.

The proof of this assertion, given in Laha, Lukacs and Newman (1960), requires more powerful tools than the proof of Theorem 4.1. Moreover, it is very likely that the assumption concerning p and n (which is always satisfied if n is sufficiently large) is superfluous and is necessitated only by the technique used in the proof. It would be of interest to modify the proof so as to avoid this restriction.

The first investigations of characterization problems dealt with specific statistics, such as the sample mean and the sample variance, and the hypothesis of their independence.* This assumption was used to determine the population distribution function, except for the numerical value of some parameters. The solution of these problems was usually carried out by deriving a differential equation which the characteristic function must satisfy and by determining those solutions of this differential equation which are characteristic functions. The decision whether a solution of the differential equation is a characteristic function is frequently the most difficult part of this procedure. It appeared therefore desirable to find general properties of characteristic functions which satisfy certain differential equations. This led then also to investigations concerning the analytic properties of characteristic functions which occur in characterization problems.

The following result, concerning differential equations is due to Zinger and Linnik (1956) and (1957). For its formulation we must introduce certain notations and definitions.

Let

$$\Sigma A_{j_1 \ldots j_n} f^{(j_1)}(t) \ldots f^{(j_n)}(t) = c[f(t)]^n \qquad \ldots \quad (4.1)$$

be an ordinary differential equation of order m. The $A_{j_1 \ldots j_n}$ are real constants and the sum is taken over all non-negative integers $j_1, \ldots, j_n$ which satisfy the relation

$$j_1 + j_2 + \cdots + j_n \leqslant p. \qquad \ldots \quad (4.2)$$

Here p is an integer such that at least one of the coefficients $A_{j_1 \ldots j_n}$ with $j_1 + \cdots + j_n = p$ is different from zero. We adjoin to the differential equation (4.1) the polynomial

$$A(x_1, \ldots, x_n) = \frac{1}{n!} \sum{}^{*} \sum A_{j_1 \ldots j_n} x_{k_1}^{j_1} \ldots x_{k_n}^{j_n} \qquad \ldots \quad (4.3)$$

where the first summation Σ^* runs over all permutations $(k_1, \ldots, k_n)$ of the first n positive integers while the second summation is taken over all $(j_1, \ldots, j_n)$ which satisfy (4.2).

The differential equation (4.1) is said to be positive definite if its adjoint polynomial (4.3) is non-negative.

Theorem 4.2 : *Suppose that the function $f(t)$ is, in a certain neighbourhood of the origin, a solution of the positive definite differential equation* (4.1) *and assume that* $m \geqslant n-1$. *If the solution is a characteristic function, then it is necessarily an entire function.*

---

* Occasionally weaker assumptions were used; these will be discussed in Sections 5 and 6.

Theorem 4.2 is a very interesting result concerning the analytic properties of the solutions of certain ordinary differential equations. Moreover, differential equations of the form (4.1) occur in many characterization problems. This is not only the case if one tries to characterize populations by the independence of a polynomial statistic and the sample mean, but also by the weaker assumption (to be discussed in Section 5) that a polynomial statistic has constant regression on the sample mean. However, in applying Theorem 4.2 one is greatly handicapped by the severe restrictions contained in its assumptions. As an example we mention the situation treated by Theorem 4.1. In this case, the assumption of Theorem 4.2 that $m \geqslant n-1$ restricts its application to samples of size $n \leqslant 3$. Similar examples can be given and it seems therefore desirable to try to weaken the assumptions of Theorem 4.2. Moreover, it would be useful (for instance in connection with characterizations of the Gamma distribution) to obtain results which use weaker assumptions but state only that the solutions of a differential equation which are characteristic functions are regular in a horizontal strip or in a half-plane.

Theorem 4.2 indicates the recent trend of the characterization problem; this trend can be described as an attempt to derive analytical properties of the characteristic functions rather than to obtain a complete determination of the population. We mention next a few results of this type.

A polynomial* $P(x_1, x_2, ..., x_n)$ of degree m is said to be *admissible* if the coefficients of the terms x_j^m $(j = 1, 2, ..., n)$ are not zero.

Theorem 4.3 : *Let $X_1, X_2, ..., X_n$ be n independently (but not necessarily identically) distributed random variables. Let $P(X_1, X_2, ..., X_n)$ and $Q(X_1, X_2, ..., X_n)$ be two admissible polynomial statistics. If P and Q are independent then each X_j $(j = 1, 2, ..., n)$ has finite moments of all orders.*

It is possible to obtain more information about the distribution functions of the random variables X_j if one assumes that one of the polynomials is a linear statistic.

Theorem 4.4 : *Let $X_1, X_2, ..., X_n$ be n independently (but not necessarily identically) distributed random variables with characteristic functions $f_1(t), f_2(t), ..., f_n(t)$ respectively. Let $P = P(X_1, X_2, ..., X_n)$ be an admissible polynomial statistic and let $L = \sum_{j=1}^{n} a_j X_j$ (with $a_j \neq 0$ for $j = 1, 2, ..., n$) be a linear form. If P and L are independent then the characteristic functions $f_j(t)$ $(j = 1, 2, ..., n)$ are entire functions of finite order.*

Theorems 4.3 and 4.4 are due to Zinger (1958).

Suppose that the conditions of Theorem 4.2 hold and that the characteristic functions $f_j(t)$ have no zeros in the entire complex plane. Then it is easy to show that the random variables X_j are normally distributed. Naturally, it is not appropriate to impose such a condition on the characteristic functions. Instead it is desirable to find a condition on the polynomial P which assures that the $f_j(t)$ have no zeros. For the case of identically distributed random variables, Linnik (1956) succeeded in finding

* We assume that similar terms have been collected in every polynomial which we consider.

such a condition. This leads to characterizations of the normal population. We remark that the characterization by the independence of k_p and $\bar{X}$ as well as the characterization by the independence of m_p and $\bar{X}$ can be obtained from these general theorems.

The proof of these results is rather complicated. We refer the reader either to the original papers by Zinger (1958), Linnik (1956) or to Laha and Lukacs (1963).

5. Regression problems

If one examines the proof of Theorem 4.1 then one notices that the assumption of the independence of the statistics s^2 and $\bar{X}$ is not fully used. This leads to the question whether the assumptions of the Theorem 4.1 could be modified so as to contain no hypotheses which are not needed in the proof. This is possible and results in characterizations of populations by regression properties of certain statistics. We give first a definition and state a lemma which is essential in these studies.

Let X and Y be two random variables and assume that the expectation $\mathcal{E}(Y)$ of Y exists. We write $\mathcal{E}(Y|X)$ for the conditional expectation of Y, given the value of X. Clearly, $\mathcal{E}(Y|X)$ is a random variable. It is also called the regression of Y on X.

We say that the random variable Y has constant regression on the random variable X if the relation.

$$\mathcal{E}(Y|X) = \mathcal{E}(Y) \quad \ldots \quad (5.1)$$

holds almost everywhere.

Lemma 5.1 : *Let X and Y be two random variables and suppose that the expectation $\mathcal{E}(Y)$ exists. Y has constant regression on X if, and only if, the relation*

$$\mathcal{E}(Ye^{itX}) = \mathcal{E}(Y)\mathcal{E}(e^{itX}) \quad \ldots \quad (5.2)$$

holds for all real t.

If one multiplies (5.1) by e^{itX} and takes the expectation then one sees immediately that (5.2) is a necessary condition and it is not difficult to prove its sufficiency.

In Section 6 we shall consider two random variables X and Y such that

(a) Y has constant regression on X

(b) $\mathcal{E}(Y) = 0$.

In this case we will say that Y has zero regression on X. According to Lemma 5.1, the random variable Y has zero regression on X if, and only if, the relation $\mathcal{E}(Ye^{itX}) = 0$ holds for all real t.

We will now discuss characterizations of populations by the assumption that a polynomial statistic has constant regression on the sample mean (or equivalently on the sum $L = X_1+X_2+\cdots+X_n$). Let $X_1, X_2, \ldots, X_n$ be n independently and identically distributed random variables with common distribution function $F(x)$ [that is $X_1 \ldots, X_n$ is a sample from a population with population distribution function $F(x)$].

Let $$P = \sum A_{j_1 \dots j_n} X_1^{j_1} \dots X_n^{j_n}$$

be a polynomial statistic, the summation is here extended over all $(j_1, \dots, j_n)$ which satisfy the inequality

$$j_1 + \dots + j_n \leqslant p.$$

Suppose that P has constant regression on L, it follows then from Lemma 5.1 that

$$\mathcal{E}(Pe^{itL}) = \sum A_{j_1 \dots j_n} \mathcal{E}(X_1^{j_1} \dots X_n^{j_n} e^{itL}) = \mathcal{E}(P)\mathcal{E}(e^{itL}). \quad \dots \quad (5.3)$$

Since $\mathcal{E}(X^j e^{itx}) = i^{-j} \frac{d^j}{dt^j} f(t) = i^{-j} f^{(j)}$ and $\mathcal{E}(e^{itL}) = [f(t)]^n$, where $f(t)$ is the characteristic function of $F(x)$, we conclude from (5.3) that

$$\sum A_{j_1 \dots j_n} f^{(j_1)} \cdots f^{(j_n)} = i^{(j_1 + \dots + j_n)} \mathcal{E}(P) [f(t)]^n. \quad \dots \quad (5.4)$$

We note that (5.4) is a differential equation for $f(t)$ wich has the form (4.1); the result of Theorem 4.2, as well as the unsolved problems mentioned in connection with it, are therefore also of interest if one wishes to characterize a population by the property that a polynomial statistic has constant regression on the sample mean.

If we choose for P the sample variance s^2, then (5.4) becomes very simple and we obtain the following result.

Theorem 5.1 : *Let $X_1, X_2, \dots, X_n$ be a sample from a population whose distribution function has finite variance σ^2. The population is normal if, and only if, the sample variance s^2 has constant regression on the sample mean $\overline{X}$.*

The assumption that a random variable has constant regression on a second variable is weaker than the assumption that they are independent. It is therefore not surprising that the property of constant regression permits to characterize populations for which no characterization by the independence of two statistics is known. We mention here only the Gamma population and state without proof the following result concerning the Poisson population.

Theorem 5.2 : *Let $X_1, X_2, \dots, X_n$ be a sample from a population with distribution function $F(x)$. Let $p \geqslant 1$ and $r \geqslant 1$ be two integers and assume that*

(i) *$F(x)$ has moments up to order $p+r$,*

(ii) *$F(x) = 0$ for $x < 0$ while $F(x) > 0$ for $x \geqslant 0$.*

The statistic $k_{p+r} - k_p$ has constant regression on $k_1 = \overline{X}$ if, and only if, $F(x)$ is a Poisson distribution.

If we omit assumption (ii), we can characterize a wider family of distributions; it can be shown that this family consists of a convolution of a Poisson distribution, the conjugate to a Poisson distribution and a normal distribution. Either of these three factors may be absent, assumption (ii) assures the absence of the normal and the conjugate Poissonian component. For details we refer the reader to Lukacs (1961). It would be interesting to obtain similar characterizations of finite convolutions of Poisson type distributions.

In the next section we will discuss in detail the characterization of a population by constant (zero) regression. Before formulating this problem we wish to make a few remarks concerning other regression problems. The characterization problems treated in this section can be extended by considering polynomial regression. The resulting computations become very tedious, some results are known concerning quadratic regression. There are also many interesting investigations concerning linear stochastic structures. A discussion of these topics would exceed the scope of this paper, some results as well as references can be found in Laha and Lukacs (1963).

6. Characterization by zero regression

In this section we discuss a characterization of the binomial and of the negative binomial population by the property that a certain polynomial statistic has zero regression on the sample mean. It is convenient to express this polynomial statistic in terms of the power sums

$$s_k = \sum_{j=1}^{n} X_j^k \qquad \ldots \quad (6.1)$$

as

$$T = ns_4+(n-4)s_3s_1+(3-2n)s_2^2+ s_2s_1^2-ns_3+(n+1)s_2s_1-s_1^3. \qquad \ldots \quad (6.2)$$

This is a symmetric, inhomogeneous statistic of degree 4 which does not seem to admit a simple representation in terms of other, more familiar functions of the observations. We formulate now our result:

Theorem 6.1 : *Let $X_1, X_2, \ldots, X_n$ be a sample from a population whose distribution function is non-degenerate and has a finite moment of the third order and denote the cumulants of order 1 and 2 of $F(x)$ by κ_1 and κ_2 respectively. The characteristic function $f(t)$ of $F(x)$ has the form*

$$f(t) = \left[\frac{\kappa_2}{\kappa_1}+\frac{\kappa_1-\kappa_2}{\kappa_1}\, e^{it}\right]^{\kappa_1^2/(\kappa_1-\kappa_2)} \qquad \ldots \quad (6.3)$$

if, and only if, the statistic T, given by (6.2), *has zero regression on $L = X_1+X_2+\ldots+X_n$.*

The distribution function of $f(t)$ is a binomial distribution if $0 < \kappa_2/\kappa_1 < 1$ and $n = \kappa_1^2/(\kappa_1-\kappa_2)$ is a positive integer; it is a negative binomial distribution if $0 < \kappa_1/\kappa_2 < 1$ and is the conjugate to a negative binomial distribution, shifted by $r = \kappa_1^2/(\kappa_2-\kappa_1)$ in case $\kappa_1 < 0$.

We write T in terms of the augmented symmetric functions (see f.i. David and Kendall (1949)) and obtain

$$T = (n-2)[3\ 1]-2(n-2)[2^2]+[2\ 1^2]+(n-2)[2\ 1]-[1^3] \qquad \ldots \quad (6.4)$$

or

$$T = (n-2)\,\Sigma\, X_i^3\, X_j-4(n-2)\,\Sigma\, X_i^2\, X_j^2+2\Sigma\, X_i^2\, X_j\, X_k$$
$$+(n-2)\,\Sigma\, X_i^2\, X_j-6\Sigma\, X_i\, X_j\, X_k \qquad \ldots \quad (6.4a)$$

where the summation goes over all subscripts i, j, k, which are different.

It follows from Lemma 5.1 that the relation

$$\mathcal{E}(T\, e^{itL}) = 0 \qquad \dots \quad (6.5)$$

holds for all real t.

Since $$\mathcal{E}X\,({}^{k}\, e^{itX}) = i^{-k} \frac{d^k}{dt^k} f(t)$$

we obtain easily from (6.5) the differential equation

$$f'''f'f^{n-2}-2(f'')^2f^{n-2}+f''(f')^2f^{n-3}+\; if\,''f'f^{n-2}-i(f')^3f^{n-3} = 0.$$

Since $f(t)$ is continuous and $f(0) = 1$, there exists a $\Delta > 0$ such that $f(t) \neq 0$ for $|t| < \Delta$. We restrict in the following t to this neighbourhood of $t = 0$ and can therefore divide the preceding equation by $[f(t)]^n$ and get

$$\frac{f'''}{f}\cdot\frac{f'}{f}-2\left(\frac{f''}{f}\right)^2+\frac{f''}{f}\left(\frac{f'}{f}\right)^2+i\frac{f''}{f}\frac{f'}{f}-i\left(\frac{f'}{f}\right)^3 = 0. \qquad \dots \quad (6.6)$$

We put $\varphi(t) = ln\, f(t)$ and obtain the differential equation

$$i\,\varphi'''\varphi'-2i(\varphi'')^2-\varphi''\varphi' = 0 \qquad (|t| < \Delta). \qquad \dots \quad (6.7)$$

The initial conditions to be satisfied are

$$\varphi(0) = 0;\ \varphi'(0) = i\kappa_1;\ \varphi''(0) = -\kappa_2. \qquad \dots \quad (6.7a)$$

We obtain, by means of a simple computation,

$$\varphi(t) = \frac{\kappa_1^2}{\kappa_1-\kappa_2}\, ln\left[\frac{\kappa_2}{\kappa_1}+\frac{\kappa_1-\kappa_2}{\kappa_1}\, e^{it}\right]$$

so that $$f(t) = \left[\frac{\kappa_2}{\kappa_1}+\frac{\kappa_1-\kappa_2}{\kappa_1}\, e^{it}\right]^{\kappa_1^2/(\kappa_1-\kappa_2)} \qquad \dots \quad (6.8)$$

If $f(t)$, as given by (6.8), is a characteristic function, then it is analytic characteristic function so that formula (6.8) is valid not only for $|t| < \Delta$ but for all real t.

We still have to find conditions which assure that $f(t)$ is a characteristic function.

We note first that $\kappa_2 = 0$ as well as $\kappa_2/\kappa_1 = 1$ and $\kappa_1 = 0$ lead to degenerate distributions, these cases are therefore excluded by the assumptions of the theorem.

In discussing (6.8) we must distinguish three cases.

Case 1 : $0 < \kappa_2/\kappa_1 < 1$. Then $\kappa_1-\kappa_2 > 0$ and $n = \kappa_1^2/(\kappa_1-\kappa_2) > 0$. We put $p = \kappa_2/\kappa_1$, $q = 1-p$ and write (6.8) as

$$f(t) = (p+qe^{it})^n. \qquad \dots \quad (6.9a)$$

If n is an integer then this is the characteristic function of the binomial distribution. If $n > 0$ is not an integer then it is easily seen that $f(t)$ is the Fourier-Stieltjes transform of a function which is not monotone. Therefore (6.9a) is a characteristic function only if n is an integer.

Case 2 : $0 < \kappa_1/\kappa_2 < 1$: Then $\kappa_2-\kappa_1 > 0$ and $r = \kappa_1^2/(\kappa_2-\kappa_1) > 0$. We put $p = \kappa_1/\kappa_2$, $q = 1-p$ and obtain from (6.8)

$$f(t) = p^r(1-qe^{it})^{-r}. \qquad \dots \quad (6.9b)$$

This is the characteristic function of the negative binomial distribution.

Case 3 : $\kappa_1 < 0$: Then $-\kappa_1 > 0$ and $\kappa_2 - \kappa_1 > 0$, we put $p = -\kappa_1/(\kappa_2-\kappa_1)$, $q = 1-p$ so that $0 < p < 1$, $0 < q < 1$ and write $r = \kappa_1^2/(\kappa_2-\kappa_1) > 0$. Using these notations we can transform (6.8) into

$$f(t) = p^r e^{-itr}(1-q\, e^{-it})^{-r}. \quad \dots \quad (6.9c)$$

We see then immediately that

$$e^{-itr}f(-t) = p^r(1-q\, e^{it})^{-r} \ ;$$

this is the last statement of the theorem.

7. Supplementary remarks

Since it is not possible to exhaust our topic without making the paper excessively long, we wish to indicate here a few problems which we did not cover in the preceding sections.

We mention first the problem of deriving conditions for the independence of certain statistics in a given population. A number of results concerning linear and quadratic statistics in normal populations may be found in Laha and Lukacs (1963).

Let X and Y be two random variables and suppose that the expectation $\mathcal{E}(X^r Y^s)$ exists. We say that X and Y are uncorrelated of order (r, s) if the relations

$$\mathcal{E}(X^i Y^j) = \mathcal{E}(X^i)\mathcal{E}(Y^j) \quad \dots \quad (7.1)$$

hold for $i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$. Two random variables uncorrelated of order (1, 1) are clearly uncorrelated in the usual sense. It is well known that in a normal population two linear statistics are independent if, and only if, they are uncorrelated [of order (1,1)]. A few results concerning the uncorrelatedness of higher order and independence of certain linear and quadratic statistics in normal variables were also obtained. Linnik (1954), (1958-59) studied also the more general problem of two polynomial statistics in a normal population and raised some open problems and gave also some interesting results. The starting point for investigations on independent statistics in the normal population is probably the famous theorem of Cochran (1933) on the decomposition of chi-squares. The method of characteristic functions can also be applied to some problems concerning stochastic processes; we mention in particular independently and identically distributed stochastic integrals; their study leads to characterizations of the Wiener-Lévy process and it might be possible to get similar results for stable processes.

References

Cochran, W. G. (1933) : The distribution of quadratic forms in a normal system with applications to the analysis of variance. *Proc. Camb. Phil. Soc.*, **30**, 178–191.

Darmois, G. (1951) : Sur une propriété caractéristique de la loi de Laplace. *C. R. Acad. Sci. Paris*, **232**, 1999–2000.

——— (1953) : Analyse générale des liaisons stochastiques. *Revue Intern. Stat. Inst.*, **21**, 2-8.

David, F. N. and Kendall, M. G. (1949) : Tables of symmetric functions, Part I. *Biometrika*, **36**, 431–449.

Fisher, R. A. (1925) : Applications of "Student's distribution". *Metron*, **5**, 90–104.

Gnedenko, B. V. (1948) : On a theorem of S. Bernstein. *Izvestiya Akad. Nauk SSSR* (*Ser. Mat.*), **12**, 97–100.

GOLDBERG, R. R. (1961) : Fourier transforms. *Cambridge Tracts in Mathematics and Mathematical Physics* No. 52, Cambridge University Press, Cambridge.

IRWIN, J. O. (1927) : On the frequency distribution of the means of samples from a population having any law of frequency with finite moments with special reference to type II. *Biometrika*, **19**, 225-239.

——— (1929) : Note on a paper published in Biometrika 19. *Biometrika*, **21**, 431-432.

———(1930) : On the frequency distribution of the mean of samples from samples of certain Pearson types. *Metron*, **8**, 51–105.

KOTLARSKI, I. (1960) : On random variables whose quotient follows the Cauchy law. *Colloquium Math.*, **7**, 277-284.

KULLBACK, S. (1934) : An application of characteristic functions to the distribution problem of statistics. *Ann. Math. Stat.*, **5**, 263–307.

——— (1936) : On certain distribution theorems of statistics. *Bull. Am. Math. Soc.*, **42**, 407–410.

LAHA, R. G. (1956) : On the stochastic independence of a homogeneous quadratic statistic and the mean. *Vestnik Leningrad Univ.*, **11**, 25–32.

——— (1959a) : On the laws of Cauchy and Gauss. *Ann. Math. Stat.*, **30**, 1165 -1174.

——— (1959b) : On a class of distribution functions where the quotient follows the Cauchy law. *Trans. Am. Math. Soc.*, **93**, 205–215.

LAHA, R. G., LUKACS E. and NEWMAN, M. (1960) : On the independence of a sample central moment and the sample mean. *Ann. Math. Stat.*, **31**, 1028-1033.

LAHA, R. G. and LUKACS, E. (1963) : *Applications of Characteristics Functions*, Charles Griffin and Company, London (publication expected 1963)

LINNIK, YU, V. (1953) : Linear forms and statistical criteria I, II. *Ukrain. Math. Žurnal*, **5**, 207–243 and 247–290.

——— (1954) : Independent and equally distributed statistics and some statistical and analytical problems connected with them. Mimeographed notes, Indian Statistical Institute, Calcutta.

——— (1956) : On polynomial statistics in connection with the analytical theory of differential equations. *Vestnik Leningrad. Univ.*, **11**, 35–48.

——— (1956) : Determining the probability distribution by a statistics distribution. *Teoriya Veroyatn. i Primen*, **1**, 466–478.

——— (1958-59) : Polynomial statistics and polynomial ideals. *Cal. Math. Soc.*, Golden Jubilee Volume, 95–98.

LUKACS, E. (1956) : Characterization of populations by properties of suitable statistics. *Proceedings of the Third Berkeley Symposium* II, 195–214. University of California Press, Berkeley.

——— (1961) : On the characterization of a family of populations which includes the Poisson population. *Annales Univ. Sci. Budapestiensis, Sectio Math.*, **3-4**, 159–175.

MARCINKIEWICZ, J. (1939) : Sur une propriété de la loi de Gauss. *Math. Ztschr*, **44**, 612–618.

MAULDON, J. G. (1956) : Characterizing properties of statistical distributions. *Quart. Jour. Math. Oxford*, (2) **7**, 155–160.

SCHULZ, R. ARENSTORFF, and MORELOCK, J. C. (1959) : The probability distribution of the product of n random varlbles. *Amer. Math. Monthly*, **66**, 95–99.

SKITOVICH, V. P. (1953) : On a property of the normal distribution. *Doklady Akad. Nauk SSSR.*, **89**, 217–219.

——— (1954) : Linear forms in independent random variables and the normal distribution. *Izvestiya Akad. Nauk SSSR* (*Ser. Mat.*), **18**, 185-200.

STECK, G. P. (1958) : A uniqueness property not enjoyed by the normal distribution. *Ann. Math. Stat.*, **29**, 604–606.

ZINGER, A. A. (1958) : Independence of quasi-polynomial statistics and analytical properties of distributions. *Teoriya Veroyatn. i Primen*, **3**, 265–284.

ZINGER, A. A. and LINNK, YU. V. (1956) : On a certain theorem of the theory of differential equations and "invariant in the mean" statistics. *Doklady Akad. Nauk USSR*, **108**, 577–580.

——— (1957) : On a class of differential equations with an application to certain questions of regression theory. *Vestnik Leningrad. Univ.*, **12**, 121–130.

Paper received : February, 1963.

CRITERIA OF ESTIMATION IN LARGE SAMPLES*

By C. RADHAKRISHNA RAO

Indian Statistical Institute

SUMMARY. The existing criteria of consistency and efficiency of estimation have been examined in the light of recent criticisms and controversies concerning them. A new criterion called *uniform first order efficiency* which is a better indicator of the performance of an estimator in statistical inference has been introduced. It is, however, pointed out that the anomaly in the earlier criterion of efficiency can be removed by considering consistent estimators which converge to a normal distribution uniformly in compacts of the parameter space. First order efficiency by itself cannot discriminate among a large number of estimation procedures. Therefore, an additional criterion called the *second order efficiency* has been introduced, which considerably restricts the class of useful estimation procedures and by which several well established estimation procedures could be eliminated in favour of the method of maximum likelihood.

1. Introduction

Estimation, as conceived by the late Sir Ronald Fisher, is one of the methodological processes by which data are analysed or reduced for purposes of drawing inferences on the unknown population from which data are observed. For instance a sample survey of consumer expenditure may provide a mass of data which by themselves are difficult to interpret. We therefore need summary figures or *estimates* which provide a fair idea of the characteristics of the population sampled and enable us to answer a variety of questions. Has the per-capita expenditure on rice increased over time and is it different in different regions? Does a given estimate reasonably agree with what is believed to be the per-capita expenditure, or with another estimate obtained by a parallel agency? No clear indication of answers to such questions would be available without computing from the data an estimate which represents the per-capita expenditure and other quantities which indicate the possible extent of error in the estimate and guide us in making judicious statements about the population. Further questions may suggest themselves after some initial questions are answered with the estimates already obtained.

There has been a tendency to consider the problem of estimation as a part of decision theory, requiring a prestated purpose for the estimate and specification of loss resulting from any given magnitude of error in the estimate. It is not, however, my view that the latter approach should be completely abandoned. There may be situations where such an approach is necessary and appropriate as in the case of acceptance procedures in industrial statistics. But in a majority of situations the framework of decision theory may not be applicable and it may be necessary to consider the problem of estimation from a wider point of view as 'extraction of information' for drawing inferences and for recording it, as a substitute for the entire data, for possible future uses.

Since estimation, however it may be viewed, involves reduction of data, it may entail some loss of information for we are interpreting the data through the

*Lecture delivered on the occasion of the presentation of Shanti Swarup Bhatnagar award for 1959.

This paper has been included in *Contributions to Statistics* presented to Professor P. C. Mahalanobis on the occasion of his 70th birthday.

estimates. The criteria for choice of estimators should then relate to minimisation of loss of information. Unfortunately, no objective measurement of information is possible and hence the difficulty in the formulation of suitable criteria. However, asymptotic theories of estimation based on the criteria of *consistency* and *efficiency* (to be referred to as v-efficiency) have been constructed and certain methods have been shown to yield estimators satisfying these criteria. It was thought that the criteria of consistency and v-efficiency ensure minimum loss of information due to estimation as the sample size increases.

These theories are not satisfactory due to three main reasons. Firstly, all the results relate to limiting properties as the sample size tends to infinity and no indication is available of their applicability to samples of sizes ordinarily met with in actual practice. Secondly, there seem to exist infinitely many procedures leading to estimators satisfying the stated criteria and no further criteria have been suggested to distinguish among them. Thirdly, the criterion of v-efficiency does not provide a satisfactory index of the performance of an estimator from the view point of statistical inference.

I have attempted to resolve these difficulties in some ways (Rao, 1960b, 1961, 1962). Firstly, the criterion of v-efficiency has been reformulated to ensure some optimum asymptotic properties of an estimator used in the place of the sample for purposes of inference. This is called *first order efficiency*. Secondly, another criterion known as *second order efficiency* has been introduced to distinguish among different procedures leading to first order efficient estimators. On the basis of the latter criterion several well-known procedures, such as the minimum chi-square, modified minimum chi-square etc., which are considered as competitors to maximum likelihood on the basis of v-efficiency, could be eliminated. The second order efficiency also provides a partial answer to the question of sample size. Correction terms of order $O(n^{-1})$ to the estimate and of order $O(n^{-2})$ to its precision have been determined for several estimation procedures.

The present paper is intended for a further discussion of first and second order efficiencies and to introduce a new concept of *uniform efficiency* which seems to be important when asymptotic theories are considered. Some new light is thrown on the use of asymptotic variance of an estimator as an index of efficiency. Further the second order efficiency is linked with terms of order (n^{-2}) in the asymptotic expansion of the variance of an estimator. Problems requiring further investigation are indicated.

In undertaking these studies I have been guided by the basic ideas contained in two fundamental papers on estimation by Fisher (1922, 1925). I wish to record my debt of gratitude to the late Sir Ronald Fisher for the encouragement I received from him when I was working under his guidance at Cambridge and during his recent visits to the Indian Statistical Institute. I also wish to thank Professor P. C. Mahalanobis, the Director of the Indian Statistical Institute for his stimulating discussions on the logic of statistical inference and the purpose of statistics to which I have been constantly exposed.

2. Consistency

The criterion of consistency is in the nature of identifying the parameter for which a statistic is said to be an estimator. This is important from the practical point of view of interpreting the estimates. There are various definitions of consistency of which the one frequently referred to in literature is probability consistency (PC).

Definition 2A : *Probability consistency* (PC). A sequence of statistics T_n is said to be consistent for a parameter θ if $T_n \to \theta$ in probability.

But one criticism of such a definition is that it places no restriction on the statistic for any given n. An alternative definition of consistency due to Fisher, called Fisher consistency (FC) seems to be more satisfactory in this respect, but somewhat restrictive in application.

Definition 2B : *Fisher consistency* (FC). A statistic $T_n = f(S_n)$, where S_n is the empirical distribution function based on n observations and f is a weakly continuous functional defined on the space of distribution functions is said to be Fisher consistent if $f(F_\theta) \equiv \theta$, where F_θ is the true distribution function from which observations are drawn.

It is easy to see that FC $\Longrightarrow$ PC and that FC refers to a restriction on the estimate for any finite n and is not just a limiting property of a sequence of statistics. But it is applicable only in situations where independent observations are drawn from a population characterised by a distribution function.

3. Efficiency

Efficiency of an estimator, which we rename as v-efficiency because it is linked with asymptotic variance, is usually defined as follows :

Definition 3A : *v-efficiency*. Consider the class $\{T_n\}$ of consistent asymptotically normal (CAN) estimators of θ, i.e., for each T_n, $n^{\frac{1}{2}}(T_n - \theta) \xrightarrow{L} N[0, v(\theta)]$. Any member of the sub-class for which $v(\theta) = 1/i(\theta)$ is said to be an efficient estimator of θ.

It was believed that for a CAN estimator, the asymptotic variance $v(\theta)$ satisfies the inequality

$$v(\theta) \geqslant \frac{1}{i(\theta)} \qquad \ldots \quad (3.1)$$

and that an estimator with the smallest $v(\theta)$ has maximum concentration round the true value in sufficiently large samples. Unfortunately, both these results are not strictly true without any restrictions on the estimating function or the mode of convergence to normality. About ten years ago Hodges (see LeCam, 1953) constructed an example to show that the result (3.1) is not true in general. Let

$$\left.\begin{aligned} T_n &= \bar{x} \quad (|\bar{x}| \geqslant n^{-1/4}) \\ &= \alpha\, \bar{x} \, (|\bar{x}| < n^{-1/4}) \end{aligned}\right\} \qquad \ldots \quad (3.2)$$

where $\bar{x}$ is the average of n observations from $N(\theta, 1)$ and α is arbitrary. It may be verified that T_n is also CAN with

$$\left.\begin{aligned} v(\theta) &= 1, \qquad \text{for} \quad \theta \neq 0 \\ &= \alpha^2, \qquad \text{for} \quad \theta = 0 \end{aligned}\right\}$$

so that the variance at $\theta = 0$ can be made arbitrarily small. Such an estimate has been termed 'super efficient.' This example throws in doubt the exact significance of v-efficiency.

Even if there is no lower bound to asymptotic variance, the question remains as to whether we should prefer the estimator T_n as defined in (3.2) to $\bar{x}$ because of smaller asymptotic variance at least at one point and equivalence elsewhere. It can be easily seen that for any given n, T_n has better concentration than $\bar{x}$, in the sense of higher probabilities for intervals enclosing the true value, only for the special values of $\theta = 0$ and a small neighbourhood of zero, and thereafter for a continuous set of θ, T_n has less concentration than $\bar{x}$. This may also be inferred by comparing the mean square errors (m.s.e.) of T_n and $\bar{x}$. For any given n the m.s.e. of T_n is smaller than that of $\bar{x}$ for θ close to zero and thereafter it stays larger, although the difference tends to zero as θ increases. It may, however, be observed that the m.s.e. in either case tends to the corresponding asymptotic value but the anomaly arises due to convergence being not uniform in the case of T_n. We shall have occasion to stress the importance of uniform convergence in a later section of this paper. An attempt to improve the concentration in the neighbourhood of a particular value of the parameter seems to have injured the performance of the estimator at other values. A general statement to this effect is proved by LeCam (1953) using bounded risk functions. Superiority as judged by asymptotic variance function need not therefore indicate greater concentration for all values of the unknown parameter even in sufficiently large samples.

Consider another super efficient estimator U_n,

$$\left.\begin{aligned} U_n &= \bar{x} && (|\bar{x}| \geqslant n^{-1/4}) \\ &= \alpha x_m && (|\bar{x}| < n^{-1/4}) \end{aligned}\right\} \qquad \dots \quad (3.3)$$

where x_m is the sample median and α is arbitrarily small. The statistics (3.2) and (3.3) have the same asymptotic variance and are therefore indistinguishable on the basis of v-efficiency. There must, however, be some difference in the performance of these two statistics, the estimator (3.3) being essentially equivalent to the sample median when $\theta = 0$.

Since there is no lower bound to the asymptotic variance of a CAN estimator, it may be thought that an improvement over $\bar{x}$ is possible by constructing a statistic T_n with a uniformly lower asymptotic variance and thereby increasing the concentration at every value of the parameter, as at $\theta = 0$ in examples (3.2) and (3.3). LeCam (1953) has demonstrated that such an improvement is not possible for any continuous interval of the parameter and the set of points with a lower asymptotic variance has to be of Lebesgue measure zero.

Can we avoid all these troubles by considering only efficient estimators in the sense of Definition 3A and not trying to improve upon the asymptotic variance $1/i(\theta)$? The following example provides an answer to this question.

Let

$$\left.\begin{aligned} W_n &= \bar{x} && (|\bar{x}| \geqslant n^{-1/4}) \\ &= \left(\frac{2}{\pi}\right)^{\frac{1}{2}} x_m && (|\bar{x}| < n^{-1/4}) \end{aligned}\right\} \qquad \dots \quad (3.4)$$

where $\bar{x}$ is the sample mean and x_m is the sample median. W_n is also CAN with the same asymptotic variance $v(\theta) = 1$ for all θ as that of $\bar{x}$. The estimator W_n is thus indistinguishable from $\bar{x}$ so far as consistency and v-efficiency are concerned. Yet for any given large n, W_n has less concentration than that of $\bar{x}$ for all values of θ.

It is no doubt true that an estimator having a higher concentration than another for every value of θ is more useful in drawing inferences on θ from an observed estimate. That such a situation is realised for an estimator compared to another for sufficiently large n cannot be judged by comparing the asymptotic variances only as shown by examples (3.2), (3.3) and (3.4). It is, however, difficult to choose between two estimators when one does not have uniformly better concentration than another without bringing in other considerations. For instance, we may have an estimator whose distribution for a particular value of θ is highly concentrated but it will be a poor discriminator between this value of θ and other values close to it if the concentration at the other values is low. To compare the estimators $\bar{x}$, T_n, U_n and W_n, we may examine one aspect of their usefulness in statistical inference e.g., the power functions of tests based on these statistics to test the hypothesis that θ has an assigned value. It may be inferred from the optimum properties possessed by $\bar{x}$, that in large samples $\bar{x}$ and T_n tend to have the same local power (Rao, 1962) whereas U_n and W_n being equivalent to the sample median when $\theta = 0$, will have a smaller local power. Since v-efficiency does not enable us to distinguish between estimators such as $\bar{x}$ or T_n and U_n or W_n we shall consider an alternative definition of efficiency (to be called first order) which appears to be more satisfactory.

Definition 3B : *First order efficiency.* A statistic T_n is said to be efficient if

$$n^{\frac{1}{2}}|(T_n-\theta)-\beta(\theta)Z_n| \xrightarrow{L} 0 \qquad \ldots \quad (3.5)$$

where $\beta(\theta)$ is a function of θ only, and $Z_n = n^{-1}[d \log P(X_n, \theta)/d\theta]$, $P(X_n, \theta)$ being the density of the observations. The condition (3.5) implies that the asymptotic correlation between T_n and Z_n is unity.

I have shown elsewhere (Rao, 1960b) that according to definition 3B, T_n is just as efficient as $\bar{x}$, although T_n is super efficient in the sense of v-efficiency and U_n and W_n are not efficient in the new sense at $\theta = 0$ although U_n and W_n are super efficient and efficient respectively in the old sense. If the efficiency of an estimator is measured by the square of its asymptotic correlation with Z_n, then U_n and W_n have the same efficiency $2/\pi < 1$, although U_n and W_n have different asymptotic variances. It is also shown (Theorem 2 in Rao, 1962) that an estimator satisfying, or efficient in the sense of Definition 3B provides a locally more powerful test of a simple hypothesis concerning θ than any other test in sufficiently large samples. Another important consequence of Definition 3B of efficiency is that the ratio of $I(T_n)$ the Fisher's information contained in the estimator T_n to I, the total information in the sample tends to unity as $n \to \infty$ (Doob, 1934; Rao, 1961).

The Definition 3B of efficiency implies that the limiting distribution of $n^{\frac{1}{2}}(T_n-\theta)$ is normal for any given θ and in large samples, any simple hypothesis on θ can be tested by using the normal approximation. But in problems of statistical inference, it is often necessary to express our preference for different values of θ, on the basis of the estimate as in the case of interval estimation, and not just examine whether a particular value is true or not. There is thus for a given n, a need to consider the whole set of distributions of the estimator for all values of θ at least in a small interval (in large samples) where different values of θ have to be distinguished. If the distributions are to be approximated by appropriate normal distributions, it seems to be a logical necessity that the convergence to normality of the chosen estimator should be uniform in compacts of θ. Under fairly general conditions the convergence to normality of $n^{\frac{1}{2}}Z_n(\theta)$ is found to be uniform in which case the desired property is assured by the following definition of uniform first order efficiency.

Definition 3C : *Uniform first order efficiency.* An estimator is said to have uniform first order efficiency if

$$n^{\frac{1}{2}}\,|T_n-\theta-Z_n(\theta)/i(\theta)|\xrightarrow{UL}0 \qquad \dots \quad (3.6)$$

in compacts of θ, where the symbol UL stands for uniform convergence in law and $i(\theta)$ is Fisher's information per observation.

It would have been more natural to define uniform first order efficiency as

$$n^{\frac{1}{2}}\,|T_n-\theta-\beta(\theta)\,Z_n(\theta)|\xrightarrow{UL}0 \qquad \dots \quad (3.7)$$

without specifying the value of $\beta(\theta)$ as in (3.6). It appears that if the condition (3.7) is satisfied for various values of $\beta(\theta)$, then it is desirable to choose an estimator for which $\beta(\theta)$ is a minimum which is shown to be $[i(\theta)]^{-\frac{1}{2}}$ in section 4 of this paper.

4. Some lemmas

Notations and assumptions. We consider only sequences of independent and identically distributed variables with probability density $p(x, \theta)$, where θ is a parameter with values in an open interval Θ. In the case of discrete variables, $p(x, \theta)$ represents the probability of x. The probability density of n observations is denoted by $P(X_n, \theta)$. The first derivative $p'(x, \theta) = dp(x, \theta)/d\theta$ exists. Let $a(x, \theta) = p'(x, \theta)/p(x, \theta)$ and

$$i(\theta) = E_\theta[a(x, \theta)]^2$$

Fisher's information per observation be continuous in θ. The following assumptions are made in the various lemmas of this section.

Assumption I : (i) $\mu(\theta_0, \theta) = E_\theta[a(x, \theta_0)] = (\theta-\theta_0)\,i(\theta_0)+o(\theta-\theta_0)$

(ii) $[\sigma(\theta_0, \theta)]^2 = V_\theta[a(x, \theta_0)] = i(\theta_0)+o(1)$

(iii) $c(\theta_0, \theta) = \mathrm{cov}_\theta\,[a(x, \theta), a(x, \theta_0)] = i(\theta_0)+o(1)$

Assumption II : $E_\theta\left|\dfrac{p'(x, \theta)}{p(x, \theta)}\right|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$ in compacts of θ.

Assumption III :

Let $\quad b(x, \theta, \theta_0) = \log [p(x, \theta)/p(x, \theta_0)]$

(i) $\quad \xi(\theta, \theta_0) = E_{\theta_0}[b(x, \theta, \theta_0)] = -\frac{(\theta-\theta_0)^2}{2} i(\theta_0)+o(\theta-\theta_0)^2$

(ii) $\quad [\eta(\theta, \theta_0)]^2 = V_{\theta_0}[b(x, \theta, \theta_0)] = (\theta-\theta_0)^2 i(\theta_0)+o(\theta-\theta_0)^2$

(iii) $\quad \zeta(\theta, \theta_0) = \operatorname{cov}_{\theta_0}[b(x, \theta, \theta_0), a(x, \theta_0)] = (\theta-\theta_0) i(\theta_0)+o(\theta-\theta_0)$

Assumption IV : (i) $\quad \int_{E_n} \frac{d^2}{d\theta^2}[P(X_n, \theta)]dv = \frac{d^2}{d\theta^2} \int_{E_n} P(X_n, \theta)dv$

for every Lebesgue measurable set E_n,

(ii) $\quad E_\theta \left|\frac{p''(x, \theta)}{p(x, \theta)}\right|$ is bounded in compacts of θ.

The Assumptions I, II, and III are not severe. Conditions may be imposed directly on the probability density to ensure them. For instance restrictions such as those imposed by Danials (1961) on the probability density will imply the conditions (i)-(iii) of Assumption I.

Lemma 1 : *Let $\beta_n(\theta)$ be the power function of any test of the hypothesis $\theta = \theta_0$ based on a sample of n independent observations, at probability level α. Then under Assumptions* I-III

$$\overline{\lim_{n\to\infty}}\, \beta(\theta_0+\delta n^{-\frac{1}{2}}) \leqslant \Phi(a-\delta i^{\frac{1}{2}}) \qquad \ldots \quad (4.1)$$

where Φ is the distribution function of $N(0, 1)$ and a is the upper α point of $N(0, 1)$.

The limit of $\beta(\theta_0+\delta n^{-\frac{1}{2}})$ when it exists is known as Pitman power of the test. Lemma 1 gives an upper bound to Pitman power under some conditions on the probability density of the observations. Two limit theorems of a different type concerning the local power of a test have been given in an earlier paper of the author (Rao, 1962).

Let

$$Z_n(\theta) = \frac{1}{n} \frac{P'(X_n,\theta)}{P(X_n, \theta)} = \frac{1}{n} \Sigma\, a(x_i, \theta)$$

$$Y_n = \frac{1}{n} \log \frac{P(X_n, \theta)}{P(X_n, \theta_0)} = \frac{1}{n} \Sigma\, b(x_i, \theta, \theta_0)$$

$$u_n(\theta) = n^{\frac{1}{2}}[Y_n-\xi(\theta, \theta_0)]/\eta(\theta, \theta_0)$$
$$v_n(\theta) = n^{\frac{1}{2}}[Y_n+\xi(\theta_0, \theta)]/\eta(\theta_0, \theta)$$
$$w_n(\theta) = n^{\frac{1}{2}} Z_n(\theta)/[i(\theta)]^{\frac{1}{2}}.$$

Under the Assumptions I, II, and III, it is easy to show that

$$\text{(i)} \quad V_{\theta_0}[u_n(\theta)-w_n(\theta_0)] \to 0 \quad \text{as} \quad \theta \to \theta_0 \qquad \ldots \quad (4.2)$$

$$\text{(ii)} \quad V_\theta[v_n(\theta) \frac{\eta(\theta_0, \theta)}{\eta(\theta, \theta_0)} - w_n(\theta)] \to 0 \quad \text{as} \quad \theta \to \theta_0 \qquad \ldots \quad (4.3)$$

$$\text{(iii)} \quad w_n(\theta) \xrightarrow{UL} N(0, 1) \text{ in compacts of } \theta. \qquad \ldots \quad (4.4)$$

The best test of the hypothesis $H_0 : \theta = \theta_0$ against the alternative $\theta_n = \theta_0+\delta n^{-\frac{1}{2}}$ is

$$u_n(\theta_n) \geqslant c_n \qquad \ldots \quad (4.5)$$

where c_n is chosen such that the size of the test $\rightarrow \alpha$ as $n \rightarrow \infty$.

$$\lim_{n\rightarrow\infty} P_{\theta_0}[u_n(\theta_n) \geqslant c_n] = \lim_{n\rightarrow\infty} P_{\theta_0}[u_n(\theta_n)-w_n(\theta_0)+w_n(\theta_0) \geqslant c_n]$$

$$= \lim_{n\rightarrow\infty} P_{\theta_0}[w_n(\theta_0) \geqslant c_n] \text{ by (4.2).}$$

Since the limiting distribution of $w_n(\theta_0)$ is $N(0, 1)$, $c_n \rightarrow a$ the upper α point of $N(0, 1)$.

The power of the test (4.5) is

$$\beta^*(\theta_n) = P_{\theta_n}(u_n(\theta_n) \geqslant c_n)$$

$$= P_{\theta_n}(u_n(\theta_n)-w_n(\theta_n)+w_n(\theta_n) \geqslant c_n$$

$$= P_{\theta_n}\left\{ v_n(\theta_n)\frac{\eta(\theta_0, \theta_n)}{\eta(\theta_n, \theta_0)}-w_n(\theta_n)+w_n(\theta_n) > c_n+\frac{n^{\frac{1}{2}}[\xi(\theta_0, \theta_n)+\xi(\theta_n, \theta_0)]}{\eta(\theta_n, \theta_0)}\right\}$$

writing $u_n(\theta_n)$ in terms of $v_n(\theta_n)$ using their definitions.

$$\lim_{n\rightarrow\infty} \beta_n^*(\theta_n) = \lim_{n\rightarrow\infty} P_{\theta_n}\left\{ w_n(\theta_n) > c_n+\frac{n^{\frac{1}{2}}[\xi(\theta_0, \theta_n)+\xi(\theta_n, \theta_0)]}{\eta(\theta_n, \theta_0)}\right\}$$

$$= \Phi(a-\delta i^{\frac{1}{2}}) \qquad \text{using (4.4) of uniform convergence,}$$

where $-\delta i^{\frac{1}{2}} = \lim_{n\rightarrow\infty} n^{\frac{1}{2}}[\xi(\theta_0, \theta_n)+\xi(\theta_n, \theta_0)]/\eta(\theta_n, \theta_0)$. The result of Lemma 1 follows by observing that $\beta_n^*(\theta) \geqslant \beta_n(\theta)$ for each θ, where $\beta_n(\theta)$ is the power of any other test.

Lemma 2 : *Let* $n^{\frac{1}{2}}(T_n-\theta)\xrightarrow{UL} N(0, [\gamma(\theta)]^2)$ *in compacts of* θ, *where* $\gamma(\theta)$ *is bounded. Then*

(i) $\gamma(\theta)$ *is continuous if the probability density* $p(x, \theta)$ *is continuous in* θ.

(ii) $[\gamma(\theta)]^2 \leqslant 1/i(\theta)$ *under Assumptions* I–III.

We use an argument similar to that of LeCam (1960) to prove (i) of lemma 2 :

If $p(x, \theta)$ is continuous in θ the distribution function $F_{\theta,n}$ of T_n is continuous in θ and consequently the characteristic function $c_n(t, \theta)$ of $U_n = n^{\frac{1}{2}}(T_n-\theta)$ is continuous in θ. Since U_n converges uniformly, $c_n(t, \theta)$ converges uniformly to $c(t, \theta)$ the characteristic function of the asymptotic distribution $N(0, [\gamma(\theta)]^2)$. But $c(t, \theta)$ is continuous. Hence $\gamma(\theta)$ is continuous in the interval of the uniform convergence of U_n.

Let us consider the test

$$\frac{n^{\frac{1}{2}}(T_n-\theta_0)}{\gamma(\theta_0)} \geqslant \lambda_n$$

of the hypothesis $\theta = \theta_0$, at a probability level α. The power of the test at θ is

$$\beta_n(\theta) = P_\theta\{n^{\frac{1}{2}}(T_n-\theta_0) > \lambda_n\gamma(\theta_0)\}$$

$$= P_\theta\left\{\frac{n^{\frac{1}{2}}(T_n-\theta)}{\gamma(\theta)} > \lambda_n\frac{\gamma(\theta_0)}{\gamma(\theta)}-\frac{n^{\frac{1}{2}}(\theta-\theta_0)}{\gamma(\theta)}\right\}.$$

Substituting $\theta = \theta_0 + \delta n^{-\frac{1}{2}}$ and observing that the convergence to normality of $n^{\frac{1}{2}}(T_n - \theta)$ is uniform in θ, we find

$$\lim_{n\to\infty} \beta_n(\theta_0 + \delta n^{-\frac{1}{2}}) = \Phi[a - \delta/\gamma(\theta_0)] \qquad \dots \quad (4.6)$$

where the argument of Φ in (4.6) is the limit of

$$\frac{\lambda_n \gamma(\theta_0)}{\gamma(\theta)} - \frac{n^{\frac{1}{2}}(\theta - \theta_0)}{\gamma(\theta)}$$

with $\theta = \theta_0 + \delta n^{-\frac{1}{2}}$, as $n \to \infty$.

It is shown in Lemma 1 under Assumptions I–III

$$\overline{\lim_{n\to\infty}}\, \beta_n(\theta_0 + \delta n^{-\frac{1}{2}}) \leqslant \Phi(a - \delta i^{\frac{1}{2}}).$$

Hence from (4.6)

$$\Phi(a - \delta/\gamma(\theta_0)) \leqslant \Phi(a - \delta i^{\frac{1}{2}})$$

or

$$a - \delta/\gamma(\theta_0) \geqslant a - \delta i^{\frac{1}{2}}$$

i.e.,

$$\gamma^2(\theta_0) \geqslant 1/i(\theta_0) \quad \text{(for any given } \theta_0\text{)}.$$

We thus see that the asymptotic variance of CUAN (consistent uniformly asymptotically normal) estimator has Fisher's lower bound $1/i(\theta)$ when the probability density satisfies some regularity conditions. It appears then that in the examples of Hodges and LeCam, super efficiency in the sense of having asymptotic variance less than $1/i(\theta)$ has been achieved at the sacrifice of uniform convergence.

Lemma 3 : *Let*

$$n^{\frac{1}{2}} \left\{ \frac{Z_n(\theta)}{[i(\theta)]^{\frac{1}{2}}} - \frac{T_n - \theta}{\gamma(\theta)} \right\} \xrightarrow{UL} 0. \qquad \dots \quad (4.7)$$

Then

(i) $n^{\frac{1}{2}}(T_n - \theta) \xrightarrow{UL} N(0, [\gamma(\theta)]^2)$ *in compacts of* θ, *where* $\gamma(\theta)$ *is continuous, under Assumption* II *and continuity of* $p(x, \theta)$, *and*

(ii) $\gamma(\theta) = [i(\theta)]^{-\frac{1}{2}}$ *under Assumptions* II *and* IV.

Under Assumption II,

$$\frac{n^{\frac{1}{2}} Z_n(\theta)}{[i(\theta)]^{\frac{1}{2}}} \xrightarrow{UL} N(0, 1) \qquad \dots \quad (4.8)$$

and hence

$$\frac{n^{\frac{1}{2}}(T_n - \theta)}{\gamma(\theta)} \xrightarrow{UL} N(0, 1) \qquad \dots \quad (4.9)$$

since by the condition (4.7) of Lemma 3, the difference of (4.8) and (4.9) $\xrightarrow{UL} 0$, Hence the result (i) of Lemma 3 follows.

Consider the test

$$n^{\frac{1}{2}}(T_n - \theta_0) \geqslant c_n \gamma(\theta_0) \qquad \dots \quad (4.10)$$

of the hypothesis $\theta = \theta_0$ at a probability level α, where $c_n \to a$, the upper α probability point of $N(0, 1)$. The power of the test (4.10) at $\theta_n = \theta_0 + \delta n^{-\frac{1}{2}}$ is

$$\beta_n(\theta_0+\delta n^{-\frac{1}{2}}) = P_{\theta_n}(n^{\frac{1}{2}}(T_n-\theta_0) \geqslant c_n\gamma(\theta_0)$$

$$= P_{\theta_n}\left\{\frac{n^{\frac{1}{2}}(T_n-\theta)}{\gamma(\theta)} \geqslant \frac{c_n\gamma(\theta_0)}{\gamma(\theta)} - \frac{n^{\frac{1}{2}}(\theta-\theta_0)}{\gamma(\theta)}\right\}$$

$$\lim_{n\to\infty} \beta_n(\theta_0+\delta n^{-\frac{1}{2}}) = \Phi[a-\delta/\gamma(\theta_0)] \quad \ldots \quad (4.11)$$

using the uniform convergence proved in (i) of Lemma 3. It appears from (4.11) that a test of the hypothesis $\theta = \theta_0$ based on T_n does not attain the full Pitman power $\Phi(a-\delta i^{\frac{1}{2}})$ unless $\gamma^2 = i^{-1}$. It is therefore interesting to know whether the condition (4.7) of Lemma 3 itself implies that $\gamma^2 = i^{-1}$. I have been able to establish this result only under the additional Assumption IV but it is worth examining whether such a strong assumption is necessary.

Under condition (i) of Assumption IV we have the expansion of the power function

$$\beta_n(\theta_0+\delta n^{-\frac{1}{2}}) = \beta_n(\theta_0) + \frac{\delta}{n^{\frac{1}{2}}}\beta_n'(\theta_0) + \frac{\delta^2}{2n}\beta_n''(\theta') \quad \ldots \quad (4.12)$$

and under condition (ii) of Assumption IV, $\beta_n''(\theta')/n$ is bounded in an interval of θ' enclosing θ_0. From (4.12) we find

$$\lim_{\delta\to 0}\lim_{n\to\infty} \frac{\beta_n(\theta_0+\delta n^{-\frac{1}{2}})-\beta(\theta_0)}{\delta} = \lim_{n\to\infty} n^{-\frac{1}{2}}\beta_n'(\theta_0).$$

Hence
$$\lim_{\delta\to 0} \frac{\Phi(a-\delta/\gamma)-\Phi(a)}{\delta} = \lim_{n\to\infty} n^{-\frac{1}{2}}\beta_n'(\theta_0) \quad \ldots \quad (4.13)$$

The limit of the R.H.S. of (4.13) is

$$(i/2\pi)^{\frac{1}{2}}e^{-a^2/2} \quad \ldots \quad (4.14)$$

using the result of Theorem 1 in an earlier paper (Rao, 1962). The value of the L.H.S. of (4.13) is

$$(2\pi\gamma^2)^{-\frac{1}{2}}e^{-a^2/2}. \quad (4.15)$$

Comparing (4.14) and (4.15) we find $\gamma^2 = i$ which establishes (ii) of Lemma 3.

Lemma 4 : *Let $\{n^{\frac{1}{2}}(T_n-\theta), n^{\frac{1}{2}}Z_n(\theta)\} \to$ in law to a bivariate normal distribution uniformly in compacts of θ, with the asymptotic covariance matrix*

$$\begin{pmatrix} \beta^2(\theta)/i(\theta) & \rho(\theta)\beta(\theta) \\ \rho(\theta)\,\beta(\theta) & i(\theta) \end{pmatrix}.$$

Then under Assumptions I–III

$$\beta(\theta) = 1 \Longrightarrow \rho(\theta) = 1 \Longrightarrow n^{\frac{1}{2}}\,|\,Z_n(\theta)-i(T_n-\theta)\,| \overset{UL}{\to} 0.$$

The Lemma 4 implies that v-efficiency of UCAN is equivalent to uniform first order efficiency.

Consider the test

$$n^{\frac{1}{2}}[T_n-\theta_0+\lambda Z_n(\theta_0)] > c_n\sigma$$

where $\sigma^2 = 1/i(\theta)+\lambda^2 i(\theta_0)+2\lambda\rho(\theta_0)$ the asymptotic variance of the test statistic. Using an argument similar to that of Lemma 3, the Pitman power of the test is

$$\Phi(a-\delta(1+\lambda i)/\sigma).$$

By the result of Lemma 1,

$$\frac{\delta(1+\lambda i)}{\sigma} \leqslant \delta i^{\frac{1}{2}}, \text{ for any arbitrary } \lambda$$

or

$$(1+\lambda i)^2 \leqslant 1+2\lambda i\rho+\lambda^2 i^2$$

which implies $\rho = 1$ at $\theta = \theta_0$ (any chosen value). The asymptotic variance of $n^{\frac{1}{2}}[Z_n(\theta)-i(T_n-\theta)]$ is then zero, and since the convergence is assumed to be uniform the desired result follows.

The results of Lemmas 1—4 under the conditions assumed on the probability density of the observations can be summarised as follows.

(i) If T_n is UCAN, the asymptotic variance of T_n has Fisher's lower bound $1/ni$. This implies that the concept of v-efficiency is not void when the class of estimators is restricted to UCAN.

It may be noted that the existence of such a lower bound to the asymptotic variance was established by Kallianpur and Rao (1955) under some conditions on the estimator such as Fisher consistency (FC) and Frechét differentiability. Recently (Kallianpur, 1963) relaxed the restriction of Frechét differentiability to a weaker form due to Volterra. Some observations on lower bound to asymptotic variance of a CAN estimator have also been made by Bahadur (1960) from a different point of view.

(ii) Uniform first order efficiency of T_n implies that it is CUAN and v-efficient.

(iii) The converse of (ii) has been established under the additional assumption that the joint asymptotic distribution of T_n and Z_n is bivariate normal and the convergence is uniform in compacts of θ.

It may be interesting to examine other conditions under which the existence of a CUAN estimator T_n with v-efficiency implies uniform first order efficiency. Restrictions on the estimator such as those imposed by Kallianpur and Rao (1955) and Kallianpur (1963) may be sufficient.

The investigations of Section 4 show that v-efficiency is a valid and useful concept if only we restrict our consideration to estimators which are consistent and uniformly asymptotically normal in compact intervals of the unknown parameter.

5. Second order efficiency

The second order efficiency is defined in earlier papers by Rao (1961, 1962) as the minimum asymptotic variance of

$$n[Z_n-\beta(T_n-\theta)-\gamma(T_n-\theta)^2] \qquad \cdots \quad (5.1)$$

when minimised with respect to γ. Under some conditions this minimum value is equivalent to the limiting value of the difference in the actual amounts of information

contained in the sample and in the statistic. It was also shown (Rao, 1961) that for the m.l. estimate the asymptotic variance of (5.1) is the least, thus establishing its highest second order efficiency.

It may be seen that the concepts of first and second order efficiencies are not explicitly linked with any loss function. It is also not important which function of θ is under estimation. We could, for instance, define first order efficiency as

$$n^{\frac{1}{2}}\,|\,Z_n-\beta[f(T_n)-f(\theta)]\,|\to 0$$

in probability for any function f admitting a continuous first derivative. Similarly the second order efficiency could be defined as the minimum asymptotic variance of

$$n(Z_n-\beta[f(T_n)-f(\theta)]-\gamma[f(T_n)-f(\theta)]^2) \qquad \ldots \quad (5.2)$$

where f admits a continuous second derivative. The expression for the minimum asymptotic variance in either case (5.1) or (5.2) would be exactly the same. Similarly if T_n is altered as

$$T_n+\frac{g(T_n)}{n}$$

where g is a smooth function, the first and second order efficiencies remain the same although from the point of view of quadratic loss function there would be difference in terms of order $(1/n^2)$. So the first and second order efficiencies as defined refer to some intrinsic properties of an estimator (statistic) used as a substitute for the whole sample for purposes of inference on the unknown parameter.

In a discussion on my paper (Rao, 1962), Lindley thought that the superiority of the m.l. estimate is probably established through some specific loss function implicit in the definition of second order efficiency. It is, therefore, proposed to compare different estimators in a more direct way by assuming a quadratic loss function. Before doing this, the procedure has to be cleared of some unpleasantness arising out of some samples of relatively small frequency leading to large deviations in the estimator and making the expected loss unduly large. We shall, therefore, omit a portion of the sample space and compare the performance of estimators over the rest of the sample space. Usually the total probability of the portion so omitted rapidly diminishes to zero as the sample size increases and the value of the estimator over this portion could be defined arbitrarily except that it should be bounded.

We shall consider the case of the finite multinomial distribution as in the earlier paper (Rao, 1961). Let us represent the theoretical frequencies in the k cells by

$$\pi_1(\theta), \ldots, \pi_k(\theta)$$

where θ is an unknown parameter, the observed proportions by

$$p_1, \ldots, p_k$$

and the estimating equation by

$$f(\theta, p_1, \ldots, p_k) = 0 \qquad \ldots \quad (5.3)$$

where

$$f(\theta, \pi_1(\theta), \ldots, \pi_k(\theta)) \equiv \theta$$

so that the estimator satisfies Fisher consistency. We shall assume that f as a function of θ, $p_1, \ldots, p_k$ admits third order partial derivatives which are bounded in a closed region P of the cube C

$$0 \leqslant p_i \leqslant 1, \qquad i = 1, \ldots, k$$

and for values of θ satisfying (5.3) with $(p_1, \ldots, p_k) \epsilon P$. The true point $\pi_1(\theta), \ldots, \pi_k(\theta)$ is assumed to be an interior point of P. Let θ^* be a solution of the equation (5.3) such that $\theta^* \to \theta$ as $p_i \to \pi_i(\theta)$. Then expanding $f(\theta^*, p_1, \ldots, p_k)$ by Taylor's theorem at θ, $\pi_1(\theta), \ldots, \pi_k(\theta)$, we have

$$\frac{\delta f}{\delta \theta}(\theta^* - \theta) + \Sigma \frac{\delta f}{\delta \pi_r}(p_r - \pi_r)$$

$$= -\frac{1}{2} \Sigma\, \Sigma \frac{\delta^2 f}{\delta \pi_r \delta \pi_s}(p_r - \pi_r)(p_s - \pi_s)$$

$$-\frac{1}{2}(\theta^* - \theta)^2 \frac{\delta^2 f}{\delta \theta^2} - (\theta^* - \theta) \Sigma \frac{\delta^2 f}{\delta \theta \delta \pi_r}(p_r - \pi_r) + \varepsilon, \qquad \ldots \quad (5.4)$$

Due to the boundedness of the third order partial derivatives, if we define θ^* arbitrarily in $C - P$, except that it should be bounded, it follows that

$$E(\theta^* - \theta) = O(n^{-1}), \qquad E(\varepsilon^2) = O(n^{-3}).$$

If the equation (5.3) is such that first order efficiency is satisfied then

$$\frac{\delta f}{\delta \pi_r} \div \frac{\delta f}{\delta \theta} = -\frac{1}{i} \frac{\pi_r'}{\pi_r}$$

as shown in (Rao, 1961) in which case, dividing (5.4) by $\delta f / \delta \theta$, the left hand side expression can be written

$$\theta^* - \theta - Z_n(\theta)/i$$

where $Z_n = \Sigma[\pi_r'(p_r - \pi_r)/\pi_r]$. If the right hand side of (5.4) without ε, divided by $\partial f / \partial \theta$ and $(\theta^* - \theta)$ replaced by Z_n/i is represented by S_n, we have the approximate relation

$$\theta^* - \theta - Z_n/i \sim S_n. \qquad \ldots \quad (5.5)$$

Hence

$$E(\theta^* - \theta) \sim E(S_n) = b(\theta)/n$$

where $b(\theta)/n$ is the bias in the estimator up to terms of $O(1/n)$. Such a bias has no effect if the mean square error is evaluated up to terms of $O(1/n)$. Otherwise correction for bias seems to be necessary. The correction can easily be done by considering the estimator

$$\hat{\theta} = \theta^* - \frac{b(\theta^*)}{n}$$

in which the bias is $o(1/n)$. We shall evaluate $E(\hat{\theta} - \theta)^2$ upto terms of $O(1/n^2)$.

Consider the approximate relationship

$$E(\theta^*) \sim \theta + \frac{b(\theta)}{n}$$

which on differentiation with respect to θ yields

$$nE(\theta^* Z_n) \sim 1 + \frac{b'(\theta)}{n}. \qquad \dots \quad (5.6)$$

Further

$$V(\hat{\theta}) \sim V(\theta^* - Z_n b'(\theta)/ni)$$
$$\sim V(\theta^*) - 2b'(\theta)/n^2 i$$

using (5.6) and

$$V(\theta^* - Z_n/i) = V(\theta^*) + V(Z_n/i) - 2 \text{ cov } (\theta^*, Z_n/i)$$
$$= V(\theta^*) + \frac{1}{ni} - \frac{2}{ni}\left(1 + \frac{b'}{n}\right)$$
$$= V(\theta^*) - \frac{2b'}{n^2 i} - \frac{1}{ni}$$
$$= V(\hat{\theta}) - \frac{1}{ni}. \qquad \dots \quad (5.7)$$

From (5.5)

$$V(\theta^* - Z_n/i) \sim V(S_n) = \frac{\psi(\theta)}{n^2} \text{ (say)}$$

Using (5.7) we have

$$V(\hat{\theta}) = \frac{1}{ni} + \frac{\psi(\theta)}{n^2} + o\left(\frac{1}{n^2}\right). \qquad \dots \quad (5.8)$$

We shall compute $\psi(\theta)$ for some methods of estimation and compare the values. The variance of θ^*, without correction for bias, is

$$V(\theta^*) = \frac{1}{ni} + \frac{\psi(\theta)}{n^2} + \frac{2b'(\theta)}{n^2 i} + o\left(\frac{1}{n^2}\right) \qquad \dots \quad (5.9)$$

(i) *Maximum likelihood.* For the method of maximum likelihood (m.l.)

$$S_n = \frac{Z_n(W_n - gZ_n)}{i^2} - \frac{\mu_{11} Z_n^2}{2i^3}$$

where

$$W_n = \Sigma(d^2 \log \pi_r/d\theta^2)(p_r - \pi_r), \; g = (\mu_{11} - \mu_{30})/i$$
$$\mu_{rs} = \Sigma \pi_i (\pi_i'/\pi_i)^r (\pi_i''/\pi_i)^s.$$

The bias in the estimator and the value of $\psi(\theta)$ are

$$\frac{b(\theta)}{n} = E(S_n) = -\frac{\mu_{11}}{2ni^2},$$

$$\psi(\theta) = n^2 V(S_n) = \frac{V[Z_n(W_n - gZ_n)]}{i^4} + \frac{\mu_{11}^2}{4i^6} V(Z_n^2)$$
$$= \frac{\mu_{02} - 2\mu_{21} + \mu_{40}}{i^3} - \frac{1}{i} - \frac{(\mu_{11} - \mu_{30})^2}{i^4} + \frac{\mu_{11}^2}{2i^4}$$
$$= \psi(\text{m.l.}). \qquad \dots \quad (5.10)$$

The variance of the m.l. estimator without correction for bias is

$$\frac{1}{ni}+\frac{\psi(\text{m.l.})}{n^2}-\frac{1}{i}\frac{d}{d\theta}\left(\frac{\mu_{11}}{ni^2}\right)+o\left(\frac{1}{n^2}\right)$$

which agrees with the expression given by Haldane and Smith (1956). It may be seen that ψ(m.l.) is connected with E_2(m.l.), the index of second order efficiency defined in the earlier paper (Rao, 1961) by the relation

$$i^2\psi(\text{m.l.}) = E_2(\text{m.l.})+\frac{\mu_{11}^2}{2i^2}\ .$$

It may be seen that the m.l. estimator corrected for bias is similar to the estimator given by Lindley (1961). For other properties of m.l. estimators reference may be made to papers by Cramér (1946), Daniels (1961), Doob (1934, 1936), LeCam (1953, 1956), Rao (1957, 1958, 1960a), Wald (1949) and others. Uniform consistency and convergence to normality of m.l. estimators are considered by Kraft (1955) and Parzen (1954).

(ii) *Minimum chi-square.* A theoretical investigation of the asymptotic properties of minimum chi-square estimates is contained in papers by Neyman (1949) and Rao (1955). The estimating equation is

$$\Sigma\, \pi_r' \frac{p_r^2}{\pi_r^2} = 0$$

and the value of S_n is

$$\left(Q+\frac{\mu_{30}}{2i^3}Z_n^2\right)+\frac{Z_n(W_n-gZ_n)}{i^2}-\frac{\mu_{11}}{2i^3}Z_n^2$$

where
$$Q = \frac{1}{2i}\Sigma\frac{\pi_r'^2}{\pi_r}(p_r-\pi_r)^2-\frac{1}{i^2}Z_n\Sigma\left(\frac{\pi_r'}{\pi_r}\right)^2(p_r-\pi_r).$$

By using the expressions already derived in Rao (1961), the bias in the minimum chi-square estimate and the value of $\psi(\theta)$ are

$$\frac{b(\theta)}{n} = \frac{1}{n}\left\{\frac{1}{2i}\Sigma\frac{\pi_r'}{\pi_r}-\frac{\mu_{30}+\mu_{11}}{2i^2}\right\}$$

$$\psi(\theta) = n^2V(S_n) = \delta+\psi(\text{m.l.})$$

where
$$\delta = \frac{1}{2i^2}\Sigma\left(\frac{\pi_r'}{\pi_r}\right)^2-\frac{\mu_{40}}{i^3}+\frac{\mu_{30}^2}{2i^4} \qquad \dots \quad (5.11)$$

which is non-negative and zero only in special cases.

(iii) *Minimum modified chi-square (Neyman, 1949).* The estimating equation is

$$\Sigma\frac{\pi_r\pi_r'}{p_r^2} = 0$$

leading to the value of S_n

$$-2\left(Q+\frac{\mu_{30}}{2i^3}Z_n^2\right)+\frac{Z_n(W_n-gZ_n)}{i^2}-\frac{\mu_{11}}{2i^3}Z_n^2.$$

The bias and $\psi(\theta)$ are

$$\frac{b(\theta)}{n} = \frac{1}{n}\left\{-\frac{1}{i}\Sigma\frac{\pi_r'}{\pi_r}+\frac{2\mu_{30}-\mu_{11}}{2i^2}\right\}$$

$$\psi(\theta) = 4\delta+\psi(\text{m.l.}).$$

(iv) *Haldane's minimum discrepancy* (*Haldane*, 1953). The estimating equation, after a slight modification which does not effect the treatment of the present paper, is

$$\Sigma\frac{\pi_r^k\pi_r'}{p_r^k} = 0$$

giving the value of S_n

$$-(k+1)\left(Q+\frac{\mu_{30}}{2i^3}Z_n^2\right)+\frac{Z_n(W_n-gZ_n)}{i^2}-\frac{\mu_{11}}{2i^3}Z_n^2,$$

The bias and $\psi(\theta)$ are

$$\frac{b(\theta)}{n}=\frac{1}{n}\left\{-\frac{(k+1)}{2i}\Sigma\frac{\pi_r'}{\pi_r}+\frac{(k+1)\mu_{30}-\mu_{11}}{2i^3}\right\}$$

$$\psi(\theta) = (k+1)^2\delta+\psi(\text{m.l.})$$

(v) *Minimum Hellinger distance.* The estimating equation is

$$\Sigma\frac{\pi_r'p_r^{\frac{1}{2}}}{\pi_r^{\frac{1}{2}}} = 0$$

giving the value of S_n

$$-\frac{1}{2}\left(Q+\frac{\mu_{30}}{2i^3}Z_n^2\right)+\frac{Z_n(W_n-gZ_n)}{i^2}-\frac{\mu_{11}}{2i^3}Z_n^2.$$

The bias and $\psi(\theta)$ are

$$\frac{b(\theta)}{n}=\frac{1}{n}\left\{-\frac{1}{2i}\Sigma\frac{\pi_r'}{\pi_r}+\frac{\mu_{30}-2\mu_{11}}{4i^2}\right\}$$

$$\psi(\theta) = \frac{\delta}{4}+\psi(\text{m.l.}).$$

(vi) *Minimum Kullback-Liebler separator.* The estimating equation is

$$\Sigma\pi_r'\log\frac{\pi_r}{p_r} = 0$$

giving the value of S_n

$$-\left(Q+\frac{\mu_{30}}{2i^3}Z_n^2\right)+\frac{Z_n(W_n-gZ_n)}{i^2}-\frac{\mu_{11}}{2i^3}Z_n^2.$$

The values of bias and $\psi(\theta)$ are

$$\frac{b(\theta)}{n}=\frac{1}{n}\left\{-\frac{1}{2i}\Sigma\frac{\pi_r'}{\pi_r}+\frac{\mu_{30}-\mu_{11}}{2i^2}\right\}$$

$$\psi(\theta) = \delta+\psi(\text{m.l.})$$

It is seen that among the six methods compared, the mean square error in the estimator corrected for bias is the least in the case of the m.l., when terms up to the order $(1/n^2)$ are considered. It may be shown more generally (following the mechanism developed in the earlier paper, Rao 1961) that under the assumptions made on the estimating equation $f(\theta, p) = 0$, the m.l. estimator has the least value for $\psi(\theta)$.

The bias and variance for estimators corrected for bias, obtained by the different methods considered in this section are given below, where δ and ψ(m.l.) are as defined in (5.10) and (5.11).

method of estimation	bias (coefficient of n^{-1})	variance of estimator corrected for bias: coefficient of n^{-1}	variance of estimator corrected for bias: coefficient of n^{-2}
maximum likelihood	$-\dfrac{\mu_{11}}{2i^2}$	$\dfrac{1}{i}$	ψ(m.l.)
minimum chi-square	$\dfrac{1}{2i}\Sigma\left(\dfrac{\pi_r'}{\pi_r}\right)-\dfrac{\mu_{30}+\mu_{11}}{2i^2}$	$\dfrac{1}{i}$	$\delta+\psi$(m.l.)
modified minimum chi-square	$-\dfrac{1}{i}\Sigma\left(\dfrac{\pi_r'}{\pi_r}\right)+\dfrac{2\mu_{30}-\mu_{11}}{2i^2}$	$\dfrac{1}{i}$	$4\delta+\psi$(m.l.)
Haldane's minimum discrepancy	$\dfrac{-(k+1)}{2i}\Sigma\left(\dfrac{\pi_r'}{\pi_r}\right)+\dfrac{(k+1)\mu_{30}-\mu_{11}}{2i^2}$	$\dfrac{1}{i}$	$(k+1)^2\delta+\psi$(m.l.)
minimum Hellinger distance	$-\dfrac{1}{2i}\Sigma\left(\dfrac{\pi_r'}{\pi_r}\right)+\dfrac{\mu_{30}-2\mu_{11}}{4i^2}$	$\dfrac{1}{i}$	$\dfrac{\delta}{4}+\psi$(m.l.)
minimum K.-L. separate	$-\dfrac{1}{2i}\Sigma\left(\dfrac{\pi_r'}{\pi_r}\right)+\dfrac{\mu_{30}-\mu_{11}}{2i^2}$	$\dfrac{1}{i}$	$\delta+\psi$(m.l.)

The expressions for bias and variance will be similar in the case of estimation of parameters in a continuous distribution. The conditions to be assumed on the estimating equation and the probability density will be very severe if an expansion of the asymptotic variance up to terms of order $(1/n^2)$ is desired. A recent paper by Linnik and Mitrafanova (1963) on the computation of the variance of the m.l. estimator in a continuous case shows the nature of the complexities involved.

REFERENCES

BAHADUR, R. R. (1960) : On asymptotic efficiency of tests and estimates. *Sankhyā*, **22**, 229–252.

CRAMÉR, H. (1946) : *Mathematical Methods of Statistics*. Princeton University Press.

DANIELS, H. E. (1961) : The asymptotic efficiency of a maximum likelihood estimator. *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 151–164.

DOOB, J. L. (1934) : Probability and Statistics. *Trans. Amer. Math. Soc.*, **36**, 759–772.

——— (1936) : Statistical estimation. *Trans. Amer. Math. Soc.*, **39**, 410–421.

FISHER, R. A. (1922) : On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy Soc.*, A, **222**, 309–365.

——— (1925) : Theory of statistical estimation. *Proc. Camb. Phil. Soc.*, **22**, 700–725.

HALDANE, J. B. S. and SMITH, SHEILA MAYNARD (1956) : The sampling distribution of a maximum likelihood estimate. *Biom.*, **43**, 96–103.

HALDANE, J. B. S. (1953) : A class of efficient estimates of a parameter. *Bull. Int. Stat. Inst.*, **33**, 231.

KALLIANPUR, G. (1963) : Von Mises functionals and maximum likelihood estimation. *Contributions to Statistics*, presented to Professor P. C. Mahalanobis on his 70th birthday.

KALLIANPUR, G. and RAO, C. R. (1955) : On Fisher's lowerbound to asymptotic variance of a consistent estimate. *Sankhyā*, **15**, 331–342.

KRAFT, CHARLES H. (1955) : Some conditions for consistency and uniform consistency of statistical procedures. *University of California Publications in Statistics*, **2**, 125–42.

LECAM, L. (1953) : On some asymptotic properties of maximum likelihood estimates and related Baye's estimates. *University of California Publications in Statistics*, **1**, 227–330.

——— (1956) : On the asymptotic theory of estimation and testing hypotheses. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, **1**, 129–156.

——— (1960) : Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, **3**, 37–98.

LINDLEY, D. V. (1961) : The use of prior probability distributions in statistical inference and decisions. *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 453-468.

LINNIK, YU. V. and MITRAFANOVA, N. M. (1963) : Some asymptotic expansions for the distribution of the maximum likelihood estimate. *Contributions to Statistics*, presented to Professor P. C. Mahalanobis on his 70th birthday.

PARZEN, E. (1954) : On uniform convergence of families of sequences of random variables. *University of California Publications in Statistics*, **2**, 23–54.

RAO, C. R. (1955) : Theory of the method of estimation by minimum chi-square. *Bull. Int. Statist. Inst.*, **35**, 25–32.

——— (1957) : Maximum likelihood estimation for multinomial distribution. *Sankhyā*, **18**, 139–148.

——— (1958) : Maximum likelihood estimation for the multinomial distribution with an infinite number of cells. *Sankhyā*, **20**, 211–218.

——— (1960a) : A study of large sample test criteria through properties of efficient estimates. *Sankhyā*, **23**, 25–40.

——— (1960b) : Apparent anomalies and irregularities in maximum likelihood estimation. *32nd Session of the Int. Stat. Inst.*, Tokyo. Reprinted with discussion in *Sankhyā*, **24**, 73–102.

——— (1961) : Asymptotic efficiency and limiting information. *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 531–546.

——— (1962) : Efficient estimates and optimum inference procedures in large samples, (with discussion). *J. Roy. Stat. Soc.*, **24**, 1, 46–72.

Wald, A. (1949) : A note on the consistency of maximum likelihood estimation. *Ann. Math. Stat.* **20**, 595–601.

SOME REMARKS ON THE POWER OF A MOST POWERFUL TEST*

By L. SCHMETTERER

Mathematical Institute, Vienna University

and

Statistical Laboratory, The Catholic University, Washington

SUMMARY. Consider test problems for a simple hypothesis against a simple alternative. It is noticed that the quotient of the power of the most powerful test and the size α of the test is a non-increasing function of α. The behaviour of this quotient near $\alpha = 0$ is studied.

Let (R, S) be a measurable space that is : R is a nonempty set and S a σ-algebra of subsets of R. Let P_0 and P_1 be two different probability measures defined over (R, S). Let α be a real number with $0 \leqslant \alpha \leqslant 1$. An S-measurable map φ_α from R into $[0, 1]$ is called a test of size α for P_0 against P_1

if
$$E(\varphi_\alpha\,; P_0) = \int_R \varphi_\alpha(x) dP_0 \leqslant \alpha.$$

$E(\varphi_\alpha\,; P_1)$ is called the power of the test. The set ϕ_α of all tests of size α is not empty. The map τ_α defined by $\tau_\alpha(x) = \alpha$ for every $x \epsilon R$ is an element of ϕ_α. τ_α is called the trivial test. For every pair P_0, P_1 and for every α there exists at least one $\bar{\varphi}_\alpha \epsilon \phi_\alpha$ with the following property :

$$E(\varphi_\alpha;\; P_1) \leqslant E(\bar{\varphi}_\alpha; P_1)$$

for all $\varphi_\alpha \epsilon \phi_\alpha$. This is nothing other than Neyman-Pearson's fundamental lemma. $\bar{\varphi}_\alpha$ is called a most powerful test for P_0 against P_1. The aim of this paper is to study the map $\alpha \rightarrow E(\bar{\varphi}_\alpha; P_1)$ for the set of all pairs (P_0, P_1) with $P_0 \neq P_1$. This last condition will not be repeated in the sequel. We will denote this map by $\alpha \rightarrow \beta(\alpha)$ whatever the pair (P_0, P_1) may be.

Lemma 1 : *For α with $0 \leqslant \alpha \leqslant 1$ we have $\alpha \leqslant \beta(\alpha) \leqslant 1$.*

For the proof it is enough to compare τ_α and $\bar{\varphi}_\alpha$ what is well known of course.

Lemma 2 : *β is nondecreasing in $0 < \alpha < 1$.*

This follows from the definition of β.

Theorem : *β is continuous in $0 < \alpha < 1$ and the map $\alpha \rightarrow \beta(\alpha)/\alpha$ is non-increasing. Furthermore, $\lim_{\alpha \rightarrow 1-0} \beta(\alpha)/\alpha = 1$ and $1 \leqslant \beta(\alpha)/\alpha \leqslant 1/\alpha$, $0 < \alpha \leqslant 1$.*

Proof : Let α_1, α_2 be real numbers with $0 < \alpha_1 \leqslant 1$, $0 < \alpha_2 \leqslant 1$. Let t_1 t_2 be positive numbers with $t_1 + t_2 = 1$. Obviously, $0 < t_1\alpha_1 + t_2\alpha_2 \leqslant 1$. Let $\bar{\varphi}_{\alpha_1}$ and $\bar{\varphi}_{\alpha_2}$ be most powerful tests. It follows that $0 \leqslant t_1\bar{\varphi}_{\alpha_1} + t_2\bar{\varphi}_{\alpha_2} \leqslant 1$ and

*This work was done with support from National Science Foundation Grant GP-96.

$E(t_1\bar{\varphi}_{\alpha_1}+t_2\bar{\varphi}_{\alpha_2} ; P_0) = t_1E(\bar{\varphi}_{\alpha_1} ; P_0)+t_2E(\bar{\varphi}_{\alpha_2} ; P_1) \leqslant t_1\alpha_1+t_2\alpha_2$ and this means that $t_1\bar{\varphi}_{\alpha_1}+t_2\bar{\varphi}_{\alpha_2}\epsilon\ \phi_{t_1\alpha_1+t_2\alpha_2}$. Therefore, we get the inequality :

$$E(t_1\bar{\varphi}_{\alpha_1}+t_2\bar{\varphi}_{\alpha_2} ; P_1) \leqslant E(\bar{\varphi}_{t_1\alpha_1+t_2\alpha_2} ; P_1)$$

or
$$t_1\beta(\alpha_1)+t_2\beta(\alpha_2) \leqslant \beta(t_1\alpha_1+t_2\alpha_2). \qquad \ldots \quad (1)$$

Lemma 2 together with (1) proves the continuity of β for $0 < \alpha < 1$. Therefore $\beta(0+0)$ exists and is $\geqslant 0$. From this and (1) we obtain the monotonicity of the function $\alpha \to \beta(\alpha)/\alpha$. It is sufficient to take $\alpha_2 = \gamma$, $t_2 = \delta/\gamma$, $0 < \delta < \gamma < 1$ (all other cases being trivial) and to make $\alpha_1 \to 0+0$ in (1).

The last two statements of the theorem are obvious. Connected with this theorem are several problems which also may be of practical interest. For instance, are there test problems, that is pairs (P_0, P_1), such that $\beta(\alpha)/\alpha = 1$ for all α ? What is the $\lim\limits_{\alpha\to 0+0} \beta(\alpha)/\alpha$ (which must exist according to the theorem) ? Can the upper bound in the inequality $\beta(\alpha)/\alpha = 1/\alpha$, $0 < \alpha < 1$, be attained ? And so on. To begin with the last problem the answer is an easy one.

Lemma 3 : *If P_0 and P_1 are orthogonal then $\beta(\alpha)/\alpha = 1/\alpha$ for $0 < \alpha \leqslant 1$.*

For the proof it is enough to consider densities f_0 resp. f_1 of P_0 resp. P_1 relative to a dominant measure μ and to define for every α the most powerful test

$$\bar{\varphi}_\alpha(x) = \begin{cases} 1 & x\,\epsilon\{x : f_1(x) > 0\} \\ \alpha & x\,\epsilon\{x : f_1(x) = 0\}. \end{cases}$$

It follows that $\beta(\alpha)/\alpha = 1/\alpha$ for $0 < \alpha \leqslant 1$.

Concerning the $\lim\limits_{\alpha\to 0+0} \beta(\alpha)/\alpha$ we would like to consider first a very well known and important example.

Example 1 : Consider two normal distributions P_0 and P_1 defined over the Borel sets of the Euclidian R_1 given by the densities $\frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$ and $\frac{1}{\sqrt{2\pi}}\, e^{-(x-a)^2/2}$ for all $x\epsilon R_1$ with $a > 0$. For every α let $k(\alpha)$ be the unique solution of $\frac{1}{\sqrt{2\pi}} \int\limits_{k(\alpha)}^{\infty} e^{-t^2/2}dt = \alpha$.

The most powerful test for P_0 against P_1 is given by

$$\bar{\varphi}_\alpha(x) = \begin{cases} 1 & k(\alpha) \leqslant x < \infty \\ 0 & -\infty < x < k(\alpha). \end{cases}$$

It follows that
$$\beta(\alpha)/\alpha = \int\limits_{k(\alpha)-a}^{\infty} e^{t^2/2}dt \Big/ \int\limits_{k(\alpha)}^{\infty} e^{t^2/2}\,dt \sim a e^{k^2(\alpha)/2}\, k(\alpha)/e^{k^2(\alpha)/2}$$

$$\sim a\sqrt{2\log\frac{1}{\alpha}}^{*} \qquad \text{for } \alpha\to 0+0.$$

Therefore
$$\lim_{\alpha\to 0+0} \beta(\alpha)/\alpha = \infty.$$

---

* If f and g are two real functions then $f(x) \sim g(x)$ for $\alpha \to a$ means $\lim f(x)/g(x) = 1$.

Another simple and important example is the following :

Example II : Let P_0 be the same distribution as in Example I and P_1 a normal distribution given by the density $\frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$ for all $x \epsilon R_1$ with $\sigma > 1$. It is easy to see that in this case $\lim_{\alpha \to 0+0} \beta(\alpha)/\alpha = \infty$ also. This leads to the conjecture that a most powerful test is always "infinitely better" than the trivial test for $\alpha \to 0+0$. But it is easy to disprove this conjecture. A most powerful test can be "almost as bad" as the trivial test.

Lemma 4 : *Let f be a density relative to Lebesgue's measure in the R_1 with the following properties : For all $x \epsilon R_1 f(x) = f(-x)$; f is continuous and strictly decreasing for $x \geqslant 0$. Let σ be a real number >1 and suppose that $x \to f(x/\sigma)/f(x)$ is strictly increasing for $x \geqslant 0$. If P_0 is given by f and P_1 by the density $x \to \frac{1}{\sigma} f(x/\sigma)$ then $\lim_{\alpha \to 0+0} \beta(\alpha)/\alpha$ is finite or infinite according as $x \to f(x/\sigma)/f(x)$ is bounded or not.*

Proof : If follows from the assumptions made that $\lim_{x \to \infty} f(x/\sigma)/f(x)$ exists. If it is finite then it is $\geqslant \sigma$. If for every α, $g(\alpha)$ is the unique solution of the equation $\int_{g(\alpha)}^{\infty} f(x)dx = \alpha/2$ then the most powerful test is given by

$$\bar{\varphi}_\alpha(x) = \begin{cases} 1 & g(\alpha) \leqslant x < \infty, -\infty < x \leqslant -g(\alpha) \\ 0 & \text{in the compliment} \end{cases}$$

It follows that $$\beta(\alpha)/\alpha = \frac{1}{\sigma} \int_{g(\alpha)}^{\infty} f(x/\sigma)dx \Big/ \int_{g(\alpha)}^{\infty} f(x)dx$$

and $$\lim_{\alpha \to 0+0} \beta(\alpha)/\alpha = \lim_{\alpha \to 0+0} \frac{1}{\sigma} f(g(\alpha)/\sigma)/f(g(\alpha))$$

and this limit is finite or infinite according as $x \to f(x/\sigma)/f(x)$ is bounded or not.

Example III : The Cauchy distribution is an example which satisfies all the assumptions of Lemma 4. Let P_0 be given by the density $x \to \frac{1}{\pi} \frac{1}{1+x^2}$ and P_1 by $x \to \frac{1}{\pi\sigma} \frac{1}{1+(x/\sigma)^2}$. Then $\lim_{x \to \infty} \frac{1}{1+(x/\sigma)^2} \Big/ \frac{1}{1+x^2} = \sigma^2$. Therefore $\lim_{\alpha \to 0+0} \beta(\alpha)/\alpha = \sigma$ and σ can be chosen as close to 1 as one wants.

In Lemmas 3 and 4 and in the Examples I, II and III the function $\alpha \to \beta(\alpha)/\alpha$ always was strictly decreasing. The conjecture that this always is true is also wrong. We consider

Example IV : Let P_0 be given by the density f_0 with

$$f_0(x) = \begin{cases} e^{-x} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

and let P_1 be given by the density f_1 with $f_1(x) = f_0(x-a)$, $a > 0$, for all $x \epsilon R_1$. It is easy to discover that a most powerful test for P_0 against P_1, if $\alpha \leqslant e^{-a}$, is given by

$$\bar{\varphi}_\alpha(x) = \begin{cases} \alpha e^a & a \leqslant x < \infty \\ 0 & -\infty < x < a. \end{cases}$$

Therefore $\quad \beta(\alpha)/\alpha = \alpha e^a/\alpha = e^a \quad$ for all, $\alpha \leqslant e^{-a}$.

The last problem to be considered is whether there exists a pair (P_0, P_1) such that the trivial test is a most powerful test. The answer is of course "no" (Lehmann, 1959). Let f_0 resp. f_1 be densities of P_0 resp. P_1 relative to a dominant measure μ. There exists an S-measurable set E of positive μ-measure such that $f_1(x) > f_0(x)$ for $x \epsilon E$. There exists also an $\varepsilon > 0$ and a set $E_1 \subset E$ such that $f_1(x) \geqslant f_0(x) + \varepsilon$ for $x \epsilon E_1$ and $\mu(E_1) \geqslant \delta > 0$. If $f_0(x) = 0$ for all $x \epsilon E_1$ except perhaps on a set of μ-measure zero, choose the test

$$\varphi_\alpha(x) = \begin{cases} 1 & x \epsilon E_1 \\ \alpha & x \epsilon R - E_1 \end{cases}$$

$0 < \alpha < 1$. Then $E(\varphi_\alpha; P_1) = \alpha P_1(R-E_1) + P_1(E_1) > \alpha$.

If $\int_{E_1} f_0(x) d\mu = \alpha_1 > 0$, choose the test

$$\varphi_{\alpha_1}(x) = \begin{cases} 1 & x \epsilon E_1 \\ 0 & x \epsilon R - E_1. \end{cases}$$

Then $E(\varphi_{\alpha_1}; P_1) = \int_{E_1} f_1(x) d\mu \geqslant \int_{E_1} f_0(x) d\mu + \varepsilon \mu(E_1) = \alpha_1 + \varepsilon\delta > \alpha_1$.

From the foregoing considerations of the lemmas and examples we have the conclusion that beyond the statements of the theorem no further general statement concerning the behaviour of $\alpha \to \beta(\alpha)/\alpha$ can be made.

Suppose that Γ is an arbitrary set of indices γ and let P_γ, $\gamma \epsilon \Gamma$, be a family of regular Radon probability measures over the Borel sets B of some locally compact space R. Let P_1 be another measure of this kind and different from all P_γ's. A test of size α, $0 \leqslant \alpha \leqslant 1$, is again a measurable map φ_α from R into $[0, 1]$ with the property

$$E(\varphi_\alpha;\ P_\gamma) \leqslant \alpha$$

for all $\gamma \epsilon \Gamma$. Suppose that the set of all measures P_γ, $\gamma \epsilon \Gamma$, and P_1 are dominated by a Radon measure μ over (R, B). Due to the weak compactness theorem (Lehmann, 1959) there always exists a most powerful test $\bar{\varphi}_\alpha$. Define β as before. It is easy to see that Lemma 1, Lemma 2 and the theorem remain true for this more general case also.

REFERENCE

LEHMANN, E. L. (1959): *Testing Statistical Hypotheses*, John Wiley-Chapman, New York-London, **67**, 354.

Paper received : December, 1962.

ON THE CONSISTENCY OF LEAST SQUARES REGRESSION

By HERMAN O. A. WOLD

University Institute of Statistics, Uppsala, Sweden

SUMMARY. In classic regression analysis there is a dualism between the treatment of a regression relation (1) in experimental and non-experimental situations, the variables $x_1, \ldots, x_h$ in the first case being regarded as a sequence of fixed numerical values, in the second as random variates having a joint probability distribution. A unified treatment of linear regression is here presented on the basis of *eo ipso* predictors, a novel name for the old notion of stochastic relations defined in terms of conditional expectations. The key theorem (Section 2) involves some rearrangement and generalization of the customary treatment.

1. Eo ipso predictors

In a stochastic relation

$$y = f(x_1, \ldots, x_h)+\nu \qquad \ldots (1)$$

with

$$E(y \mid x_1, \ldots, x_h) = f(x_1, \ldots, x_h) \qquad \ldots (2)$$

the function $f(\cdot)$ is called an *eo ipso predictor* of y.

We shall consider the case when $f(x_1, \ldots, x_h)$ is linear, say

$$y = \beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_hx_h+\nu \qquad \ldots (3)$$

with

$$E(y \mid x_1, \ldots, x_h) = \beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_hx_h. \qquad \ldots (4)$$

Writing

$$x_0 = y$$

we denote the observed values of the variables by

$$x_{0a}, x_{1a}, \ldots, x_{ha} \qquad (a = 1, \ldots, n). \qquad \ldots (6)$$

For the observed first and second order moments we write

$$m_i = \sum_{a=1}^{n} x_{ia} \qquad m_{ik} = \sum_{a=1}^{n} x_{ia}x_{ka} \qquad (i, k = 0, 1, \ldots, h). \qquad \ldots (7)$$

The theoretical moments of first and second order will be denoted

$$\mu_i = E(x_i); \qquad \mu_{ik} = E(x_ix_k) \qquad (i, k = 0, 1, \ldots, h). \qquad \ldots (8a\text{-}b)$$

Formulas (8a-b) cover the situation—typical for the analysis of nonexperimental data—when

$$y(= x_0), x_1, \ldots, x_h \qquad \ldots (9)$$

are random variates with a joint probability distribution. According to a fundamental theorem by Kolmogorov (1933), μ_0 can in this case be expressed in terms of (4),

$$\mu_0 = E(y) = E_{x_1, \ldots, x_h}[E(y \mid x_1, \ldots, x_h)] \qquad \ldots (10)$$

and similarly for $\mu_{01}, \ldots, \mu_{0h}$.

The following situation—typical for the analysis of controlled experiments—is also covered by formulas (8a-b), namely when $y(=x_0)$ is a random variate, whereas

$$x_{1a}, \ldots, x_{ha} \qquad (a = 1, \ldots, n) \qquad \ldots \quad (11)$$

are arbitrarily fixed numbers. We shall say that $x_1, \ldots, x_h$ then have a *controlled distribution.* In a factorial experiment, for example, each x_i may have two equally weighted levels; this gives a controlled distribution formed by 2^h sequences of type (11), all having the "probability" 2^{-h}. Formulas of type (10) are valid also in this case.

The two situations mentioned cover the cases usually treated in the textbooks, but they do not exhaust all possible specifications of (8a-b).

2. Least squares estimation of eo ipso predictors

As is clear from (1)–(4), the notion of *eo ipso* predictor is more or less synonymous with the notion of regression relation. The similarity extends to the procedures of parameter estimation, inasmuch as *eo ipso* predictors under very general assumptions can be consistently estimated by least squares regression.

Theorem : *Let* (3)–(4) *be a linear eo ipso predictor that satisfies the following two assumptions.*

(a) *The ergodicity assumption* :

$$\text{prob} \lim_{n\to\infty} m_i = \mu_i; \quad \text{prob} \lim_{n\to\infty} m_{ik} = \mu_{ik}; \quad (i, k = 0, 1, \ldots, h) \qquad \ldots \quad (12)$$

or in words, the observed first and second order moments have the corresponding theoretical moments as stochastic limits as the number of observations increases indefinitely;

(b) *The nonsingularity assumption* :

$$det[\mu_{ik}] > 0 \; ; \qquad (i, k = 1, \ldots, h) \qquad \ldots \quad (13)$$

or equivalently, the joint distribution of $x_1, \ldots, x_h$ involves no collinearity.

Then the least squares regression of y on $x_1, \ldots, x_h$, say

$$y = b_0 + b_1x_1 + \ldots + b_hx_h + u \qquad \ldots \quad (14)$$

will provide consistent estimates for the parameters; in symbols

$$\text{prob} \lim_{n\to\infty} b_i = \beta_i \quad (i = 0, 1, \ldots, h). \qquad \ldots \quad (15)$$

The proof of the theorem will be arranged in four steps.

(i) The assumptions (3)–(4) imply that the residual ν has zero mean and zero cross product moment with each of the variables $x_1, \ldots, x_h$; in symbols,

$$E(\nu) = 0; \; E(\nu x_i) = 0 \quad (i = 1, \ldots, h). \qquad \ldots \quad (16a\text{-}b)$$

To see this we note that (3) and (4) give

$$E(\nu \mid x_1, \ldots, x_h) = 0; \quad E(\nu \,.\, x_i \mid x_1, \ldots, x_h) = 0 \qquad \ldots \quad (17a\text{-}b)$$

Next, (17a) gives

$$E(\nu) = \int_{R_h} E(\nu \mid x_1, \ldots, x_h) dP = 0 \qquad \ldots \quad (18)$$

where the integration is taken over the joint distribution of $x_1, \ldots, x_h$. This verifies (16a), and using (17b) the same argument gives (16b).

(ii) The assumptions (3)–(4) imply that the parameters β_i satisfy the following system of linear relations,

$$\begin{cases} \beta_0 + \mu_1\beta_1 + \ldots + \mu_h\beta_h = \mu_0 \\ \mu_1\beta_0 + \mu_{11}\beta_1 + \ldots + \mu_{1h}\beta_h = \mu_{01} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \mu_h\beta_0 + \mu_{1h}\beta_1 + \ldots + \mu_{hh}\beta_h = \mu_{0h} \end{cases} \qquad \ldots \quad (19)$$

To prove, for example, the relation with μ_{0i} in the right-hand member, we multiply (3) by x_i,

$$x_i\beta_0 + x_1x_i\beta_1 + \cdots + x_hx_i\beta_h + \nu x_i = yx_i.$$

Taking expectations on both sides, the term that involves νx_i will vanish according to (16b), giving

$$\beta_0\mu_i + \beta_1\mu_{1i} + \ldots + \beta_h\mu_{hi} = \mu_{0i}.$$

(iii) The least squares regression (14), by definition, is formed so as to give the residual u the smallest possible variance. This is the same as to make the square sum

$$S = \Sigma\,(b_0 + b_1x_1 + \cdots + b_hx_h - u)^2 \qquad \ldots \quad (20)$$

as small as possible. Hence $b_0, b_1, \ldots, b_h$ should satisfy the conditions

$$\frac{dS}{db_i} = 0 \qquad (i = 0, 1, \ldots, h). \qquad \ldots \quad (21)$$

This is the same as

$$\begin{cases} b_0 + m_1b_1 + \ldots + m_hb_h = m_0 \\ m_1b_0 + m_{11}b_1 + \ldots + m_{1h}b_h = m_{10} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ m_hb_0 + m_{h1}b_1 + \ldots + m_{hh}b_h = m_{h0} \end{cases} \qquad \ldots \quad (22)$$

where the moments m_i, m_{ik} are given by (7).

(iv) The equation systems (19) and (22) are formally the same as the *normal equations* for theoretical regression coefficients β_i and empirical regressions coefficients b_i, respectively.

To conclude the argument, we infer from (13) that the solutions β_i of system (19) are uniquely determined and are continuous functions of the moments μ_i, μ_{ik}; hence making use of (12), we obtain (15). In conformity with the notations adopted, the stochastic convergence in (12) and (15) may be taken to be *convergence in probability*. The argument is valid also for other modes of convergence that are preserved under continuous transformations, for example convergence in the sense of the *strong law of large numbers* (again see Kolmogorov (1933)).

It will be noted that the solutions β_i and b_i of systems (19) and (22) as well as the corresponding residuals and residual variances are given by the same determinant expressions as in the traditional treatment of least squares regression (see, for example, Wold and Juréen (1952-53)).

3. Comments

In more or less general form the above theorem is standard material in statistical text books ; see Cramér (1945-46) and Plackett (1960). We have given the proof in detail because it differs from the customary treatment in several respects. For one thing, the theoretical normal equations (19) are derived from the assumption that the theoretical relation (3) constitutes a conditional expectation plus a residual, whereas the normal equations (22) of the sample are derived from the criterion that the observed residual should have the smallest possible variance. In this way the key of the theorem is the assumption (12). As is well known (see Cramér (1945-46) and Wold and Juréen (1952-53)) this assumption is fulfilled in a variety of situations, notably (a) the case of independent replications of a controlled experiment, and (b) the case when the variables $y, x_1, \ldots, x_h$ are given as time series that are stationary and ergodic.

The novel features of the theorem involve some slight generalization. The main point is perhaps that residual noncorrelations

$$\rho(\nu, x_i) = 0 \qquad (i = 1, \ldots, h) \qquad \ldots \quad (23)$$

usually are adopted as assumptions, while here, as seen from (16), they are obtained as implications of the basic assumption (4). Further it will be noted that independence between the residuals is not required. Thus in the case of stationary time series, for example, autocorrelation is permitted in the residuals and in the variates $x_1, \ldots, x_h$; see Wold and Juréen (1952-53) where the analysis covers the case when all variates $x_1, \ldots, x_h$ are *exogenous*, and Lyttkens (1963), where one or more of these variates are allowed to be *endogenous*, that is, are specified as the variate y subject to a lag.

As to the rationale of the basic assumption (2), we note that in applied regression analysis the approach (2) is entirely in line with the use of the relation (1) for forecasting purposes. To put it otherwise, the weak point of the customary assumptions (23) is that in most nonexperimental situations little or nothing known about the residuals and their correlation properties, whereas once we have decided for what forecasting purpose we want to use a regression relation, the corresponding assumption (2) is required as a stochastic rationale for the forecasting procedure.

The theorem has been stated without proof in earlier papers by the author (Wold (1959-60)–(1963)). The naming of the relations (1)–(2) has been changed from *unbiased* predictors (Wold, 1961) to *eo ipso* predictors (Wold, 1963). The idea of using conditional expectations in connection with forecasting problems is, of course, of old standing. Specific reference is made to the literature on interdependent systems (see Haavelmo (1944), and Hurwicz (1950) and many later works) where emphasis is placed on the specification of forecasting relations in terms of conditional expectations, but the analysis is given another twist by assuming that the variates do not have the same structural specification in the past and the future.

References

Cramér, H. (1945-46): *Mathematical Methods of Statistics.* Almqvist et Wiksells, Uppsala, N.J. University Press, Princeton.

Haavelmo, T. (1944): The probability approach in econometrics. *Econometrica*, **12**, Suppl.

Hurwicz, L. (1950): Prediction and least squares. *Statistical Inference in Dynamic Economic Models*, edited by T. C. Koopmans; 266–300, Wiley and Sons, New York.

Kolmogorov, A. (1933): Grundbegriffe der Wahrscheinlichkeitsrechnung. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, **2**, No. 3.

Lyttkens, E. (1963): Standard errors of regression coefficients in the case of autocorrelated residuals. *Time Series Analysis Symposium*, edited by M. Rosenblatt; 38–60, Wiley and Sons, New York.

Plackett, R. L. (1960): *Principles of Regression Analysis*, Oxford, Clarendon.

Wold, H. and Juréen, L. (1952-53): *Demand Analysis. A Study in Econometrics*, Almqvist et Wiksells, Stockholm; Wiley and Sons, New York.

Wold, H. (1959-60): Ends and means on econometric model building. Basic considerations reviewed. *Probability and Statistics. The Harald Cramér Volume*, edited by U. Grenander; 355–434, Almqvist et Wiksells; Stockholm, Wiley and Sons, New York.

——— (1960): A generalization of causal chain models. *Econometrica*, **28**, 443–463.

——— (1961): Unbiased predictors. *Fourth Berkeley Symposium of Statistics*, **1**, California University Press, Berkeley, 719–761.

——— (1963): Forecasting by the chain principle. *Time Series Analysis Symposium*, edited by M. Rosenblatt; 471–497, Wiley and Sons, New York.

Paper received : January, 1963.

ADDENDUM

Additional Note to the Paper "On Asymptotic Expansions For Sums Of Independent Random Variables With A Limiting Stable Distribution." By Harald Cramér, *Sankhyā*, Series A, **25**, 13–24.

While the above-mentioned paper was being printed, my attention was drawn to the paper "On the magnitude of the error in the approach to stable distributions" by M. Lipschutz, *Proc. Koninkl. Nederl. Akad. Wetensch.*, A, **59**, No. 3 (1956). In this paper, the asymptotic properties of the distribution function $F_n(x)$ are studied for the case when the independent random variables x_1, x_2 ..., are positive and have a common absolutely continuous distribution function $F(x)$.

BIOMETRIKA

Volume 50, Parts 1 and 2 **June, 1963**

CONTENTS

The subscription, *payable in advance*, is 54/- (or $8.00), per volume (including postage). Cheques should be made payable to *Biometrika*, crossed "*A/c Biometrika Trust*" and sent to the Secretary, Biometrika Office, Department of Statistics, University College, London, W. C. 1. All foreign cheques must be drawn on a Bank having a London agency.

Issued by THE BIOMETRIKA OFFICE, University College, London.

JOURNAL OF THE ROYAL STATISTICAL SOCIETY

Series B (Methodological)

40*s* each Part. Annual Subscription £3.2*s* (US $9). Post free

CONTENTS OF VOL. 25, PART 2, 1963

Obtainable through any bookseller or direct from the

ROYAL STATISTICAL SOCIETY

21 Bentinck Street, London, W. 1.

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Editors: P. C. MAHALANOBIS, C. R. RAO

SERIES A, VOL. 25 SEPTEMBER 1963 PART 3

GENERALIZATION OF THE FISHER-DARMOIS-KOOPMAN-PITMAN THEOREM ON SUFFICIENT STATISTICS

By EDWARD W. BARANKIN[1] and ASHOK P. MAITRA[2]

University of California, Berkeley

SUMMARY. The well-known and now classical theorem (or, rather, theorems) referred to in the title shows that, for a family of n-dimensional densities of product form, with *identical* 1-dimensional factor densities, the existence of a sufficient statistic of dimension $< n$ is essentially equivalent to the condition that the 1-dimensional factor density involved be of exponential type (see below for more precise descriptions). In this article we drop the feature that the factor densities be identical, and we obtain theorems again relating the existence of lower dimensional sufficient statistics with the fact of exponential type for the factor densities. Results of the classical type fall out as corollaries.

1. INTRODUCTION

In independent endeavours directed toward the elaboration of the notion of sufficiency of statistics that had been introduced by Fisher (1922), the authors Darmois (1935), Koopman (1936) and Pitman (1936)—subsequent to Fisher's (1934) own less explicit indication of the result—established the following fact, loosely stated : if $p_0(\xi, \theta)$ is a family of distribution densities on the ξ-axis, with $\theta \equiv (\theta_1, \theta_2, \ldots, \theta_\nu)$ a ν-dimensional parameter point, then the family of n-dimensional distribution densities $\prod_{i=1}^{n} p_0(x_i, \theta)$ has a sufficient statistic of dimension $s < n$ if and only if p_0 is of exponential type, having a logarithm of the form (where $x \equiv (x_1, x_2, \ldots, x_n)$)

$$b_0(\theta)+\psi_0(x)+\sum_{\lambda=1}^{r} b_\lambda(\theta)\psi_\lambda(x) \qquad \ldots \quad (1.1)$$

with $r \leqslant s$. The cited works have somewhat different accents. Darmois alone refers to the generalization of the result to the case in which ξ is a point in a Euclidean space of dimension > 1. On the other hand, Darmois takes $s = \nu$ supposing $\nu < n$ and states the result with $r = s$, whereas Koopman, who also takes $s = \nu$, concerns himself about the smallest possible r in the representation (1.1), and is able, in the case of

[1] This paper was prepared with the partial support of the Office of Naval Research, Contract Nonr-222-43. This paper in whole or in part may be reproduced for any purpose of the United States Government.

[2] This paper was prepared with the partial support of the Office of Ordnance Research, U.S.A. grant (DA-ARO(D)-31-124-G-83).

this smallest r, to relate the functions ψ_λ, $\lambda = 1, 2, \ldots, r$ to any sufficient statistic of dimension s (see Corollary 5.2 here below). Pitman does not confine himself to $s = \nu$, but his representation, like that of Darmois, has $r = s$. Pitman also is the only one to consider the case in which the carrier of the density $p_0(\cdot, \theta)$ varies with θ. The work of Koopman is distinguished by a fully precise formulation of theorems and rigor of proof : he is explicit about the assumption of analyticity of the function p_0.

The present paper is motivated by the question : to what extent can we generalize the form $\prod_{i=1}^{n} p_0(x_i, \theta)$ of the family of distribution densities in Euclidean n-space and still find that the existence of a sufficient statistic of dimension $s < n$ is equivalent to an assertion of exponential type regarding the component distributions involved? We shall show here that such theorems are obtainable when the factor densities in the n-fold product above are not necessarily identical. The Darmois-Koopman-Pitman result thus comes out as a special case.

We shall be concerned, then, with a family $\mathscr{P}_\Pi$ of probability distributions in Euclidean n-space that is specifically structured as follows. Let $\Omega_1, \Omega_2, \ldots, \Omega_n$ be n 1-dimensional open sets, and set $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$. Let Θ be a ν-dimensional open set. For each $\theta \in \Theta$ and each $i = 1, 2, \ldots, n$, let $p_i(\cdot, \theta)$ be a strictly positive probability density with respect to Lebesgue measure on Ω_i; and for each $i = 1, 2, \ldots, n$, let the function p_i, of the variables $\xi \in \Omega_i$ and $\theta \equiv (\theta_1, \theta_2, \ldots, \theta_\nu) \in \Theta$ be continuous and have continuous partial derivatives $\frac{\partial p_i}{\partial \xi}; \frac{\partial p_i}{\partial \theta_j}$, $j = 1, 2, \ldots, \nu$; $\frac{\partial^2 p_i}{\partial \xi \partial \theta_j} = \frac{\partial^2 p_i}{\partial \theta_j \partial \xi}$, $j = 1, 2, \ldots, \nu$, throughout $\Omega_i \times \Theta$. For each $x \equiv (x_1, x_2, \ldots, x_n) \in \Omega$ and $\theta \in \Theta$, let $p(x, \theta)$ be defined by

$$p(x, \theta) = \prod_{i=1}^{n} p_i(x_i, \theta). \qquad \ldots \quad (1.2)$$

Then, our family $\mathscr{P}_\Pi$ is that having the function p of (1.2) as one determination of its family density function relative to Lebesgue measure on Ω. Notice that $p(\cdot, \theta)$ is strictly positive throughout Ω for each $\theta \in \Theta$, so that the densities of $\mathscr{P}_\Pi$ all have the common carrier Ω. Our present study does not, therefore, cover the case of a variable carrier—the case whose investigation was initiated by Pitman. Indeed, our methods here are drawn from the fundamental paper of Barankin and Katz (1959) and the dimensionality problem for a variable carrier was not considered there; this problem remains to be treated.

We have not assumed analyticity of the densities p_i to start with, merely the continuous differentiability detailed above. With this lesser assumption we are able first to prove the *local* results of the type announced; this we do in Section 3. Theorem 3.1 gives, for a family $\mathscr{P}_\Pi$ defined by (1.2), necessary and sufficient conditions for the existence of a locally sufficient statistic which is locally Euclidean of dimension $s < n$ and locally continuously differentiable. Theorem 3.2 considers that there is given a statistic T which is Euclidean of dimension $s < n$ at a particular regular point x^0, and continuously differentiable about x^0; and it states necessary and sufficient conditions that T be sufficient for $\mathscr{P}_\Pi$ about x^0. Thus, the first theorem is purely

existential, while the second theorem addresses itself to a particular statistic in hand. The latter point of view is that taken in Koopman's (1936) investigation. The remainder of Section 3 presents two corollaries, which are the specializations of the two theorems to the classical case of identical factor densities.

Our ultimate aim is to obtain *global* theorems corresponding to Theorems 3.1 and 3.2, after invoking the additional hypothesis that the factor density functions are (partially or fully) analytic—the effectuating assumption made by Koopman (1936). Looking to this, we first devote Section 4 to deriving certain consequences of the analyticity hypothesis; we do this in general, not restricting ourselves to product densities. We are then ready to achieve the desired results in Section 5. Now, the analyticity hypothesis is not, in itself, enough to secure the extension of the necessary *and* sufficient conditions in the local case (Section 3) to necessary *and* sufficient conditions in the global case; Theorem 4.1(iv) provides a condition that we are obliged to use in the sufficiency part of our arguments. Consequently, in Section 5 we obtain two global theorems corresponding to the single local Theorem 3.1 : Theorem 5.1 gives necessary conditions that there exists a statistic which is sufficient for $\mathscr{P}_{\text{II}}$ in Ω' (the domain of analyticity of p; in general, a subset of Ω), is Euclidean of dimension $s < n$ in Ω', and is continuously differentiable about some regular point of Ω'; and Theorem 5.2 gives sufficient conditions. In a similar way, corresponding to the local Theorem 3.2 there are Theorems 5.4 and 5.5, the first of these giving global necessary conditions and the second giving global sufficient conditions.

Corollaries 5.1 and 5.2 are the global correspondents of the local Corollaries 3.1 and 3.2, covering the specialization of our general product-density results to the case of identical factors. The interesting fact is that the condition alluded to above, that presented in (iv) of Theorem 4.1, is automatically satisfied in the case of identical factor densities, and therefore the specializations of Theorems 5.1 and 5.2 to the identical factor case fall together into a single statement of necessary *and* sufficient conditions, this being Corollary 5.1. And likewise does Corollary 5.2 give necessary *and* sufficient conditions in a single statement on specializing Theorems 5.4 and 5.5 to the case of identical factors.

Corollary 5.2 presents the classical Koopman (1936) results. However, our context is somewhat different from that of Koopman : our definition (the modern one) of a sufficient statistic is less stringent than that of Koopman, and the statistics we deal with are not simply continuous, as in Koopman's treatment. The net effect of this is that we obtain, in Corollary 5.2, a single statement of necessary *and* sufficient conditions, whereas Koopman was not able to achieve this in his context.

Theorem 5.3 gives an alternative set of sufficient conditions in the existence situation.

In Section 2 we have recapitulated the definitions and results of our past work, on which the present investigation relies. In doing so, we have modified certain statements of results to make them more readily utilizable here. Thus, Section 2 renders the present article fairly well self-contained, and in addition will serve to shed some integrative light on our previous articles.

2. Résumé of past results

The definitions and theorems we shall need in order to achieve the conclusions at which this article is aimed have been successively built up through the four papers (Barankin and Katz, 1959 and Barankin, 1960a, 1960b, 1961). In view of this dispersion of sources, and also because certain modifications of presentation are indicated for purposes of this present work, it is advisable to devote this section to a brief, organized setting out of previous results. We shall then be in the favourable position of having to refer subsequently only to this Section 2, thereby eliminating for the reader the necessity of collating statements out of different past works.

Let $\mathcal{P}$ be a family of probability measures on the Lebesgue subsets of an open set Ω in a Euclidean n-space of points $x = (x_1, x_2, \ldots, x_n)$. The family is parametrized by a point $\theta = (\theta_1, \theta_2, \ldots, \theta_\nu)$ ranging over an open set Θ in a Euclidean ν-space. Each measure in $\mathcal{P}$ is absolutely continuous with respect to Lebesgue measure, and we consider that there is a determination, p, of the family density function which is strictly positive throughout $\Omega \times \Theta$, and is such that all the derivatives $\frac{\partial p}{\partial x_i}, \frac{\partial p}{\partial \theta_j}, \frac{\partial^2 p}{\partial x_i \partial \theta_j}, \frac{\partial^2 p}{\partial \theta_j \partial x_i}$, $i = 1, 2, \ldots, n; j = 1, 2, \ldots, \nu$, are continuous in $\Omega \times \Theta$. (From the continuity of the mixed second derivatives it follows that the order of differentiation is irrelevant).

A function T on Ω is said to be *Euclidean of dimension* r at x^0 if there is a neighbourhood of x^0 such that T maps this neighbourhood into a Euclidean r-space. Similarly, we shall speak of T being Euclidean of dimension r in a given subset of Ω, or in all of Ω.

Let T be Euclidean of dimension r at x^0; specifically, let

$$T(x) = (h_1(x), h_2(x), \ldots, h_r(x)) \qquad \ldots \quad (2.1)$$

in some neighbourhood of x^0. If the functions h_i, $i = 1, 2, \ldots, r$ are continuously differentiable (in some neighbourhood of x^0) then we say that T is *continuously differentiable about* x^0.

Let T be Euclidean of dimension r at x^0, being given by (2.1) in some neighbourhood of x^0. If T is continuously differentiable about x^0, and the Jacobian matrix

$$\left\| \frac{\partial h_i}{\partial x_j} \right\|_{i = 1, 2, \ldots, r;\; j = 1, 2, \ldots, n} \qquad \ldots \quad (2.2)$$

is of rank r at x^0, then we say that *T is regular at x^0*.

The following lemma (stated as Lemma 1.1 in Barankin and Katz (1959)) is the precise form (given by Bahadur (1954)) of the factorization theorem for sufficient statistics, and provides the referent criterion for all our recent work.

Lemma 2.1 : *A necessary and sufficient condition that the statistic T of the family $\mathcal{P}$ be a sufficient statistic for $\mathcal{P}$ is that there exists a nonnegative function f on*

$\mathcal{R}_T \times \Theta$, *and a nonnegative function* g *on* Ω, *such that* (i) *for each* $\theta \in \Theta, f(T(\cdot), \theta)$ *is Lebesgue measurable* (ii) g *is Lebesgue measurable and* (iii) *for each* $\theta \in \Theta$, *the equality*

$$p(x, \theta) = f(T(x), \theta)\, g(x) \qquad \dots \quad (2.3)$$

holds for almost all (Lebesgue) $x \in \Omega$.

In this statement, $\mathcal{R}_T$ denotes the range of the statistic T. We shall use also the notation $\mathcal{R}_{T|B}$ to denote the range of the statistic T restricted to the domain $B \subseteq \Omega$.

For a given measurable subset B of Ω, we may form from $\mathcal{R}$ the associated family of conditional probability distributions relative to B. To this derived family Lemma 2.1 may be applied; and this will result in a factorization condition of the form (2.3) restricted to B. Thus, global sufficiency for $\mathcal{P}$—that is, sufficiency over all of Ω—relates to sufficiency for derived conditional distribution families in the simple fashion of reduced domain of validity of (2.3). It follows that for discussions concerning the building up of globally sufficient statistics from functions which verify (2.3) on restricted domains, we could elect to speak in terms of sufficiency for derived conditional distribution families. But in fact there is no particular advantage for us in doing this here, and we shall therefore take the simpler route, as in Barankin (1960b), of defining restricted sufficiency directly in terms of the factorization criterion.

Definition 2.1 : A statistic T of $\mathcal{P}$ is said to be sufficient for $\mathcal{P}$ in B if there is a nonnegative function f on $\mathcal{R}_{T|B} \times \Theta$, and a nonnegative function g on B, such that (i) for each $\theta \in \Theta$, $f(T(\cdot), \theta)$ is a Lebesgue measurable function on B, (ii) g is Lebesgue measurable on B, and (iii) for each $\theta \in \Theta$, the equality

$$p(x, \theta) = f(T(x), \theta) g(x) \qquad \dots \quad (2.4)$$

holds for almost all (Lebesgue) $x \in B$.

It is handy to make also the following definition :

Definition 2.2 : A statistic T of $\mathcal{P}$ is said to be sufficient for $\mathcal{P}$ about the point $x^0 \in \Omega$ if it is sufficient for $\mathcal{P}$ in some neighbourhood of x^0.

The following lemma has been basic to all our work; it is Lemma 2.3 given by Barankin and Katz (1959), but we quote here the revised form of it stated as Lemma 2.2 in Barankin (1960b).

Lemma 2.2 : *Let* T *be a statistic of* $\mathcal{P}$ *which is Euclidean of dimension* r *at* x^0 and regular at x^0, *and which is sufficient for* $\mathcal{P}$ *about* x^0.

Then there is a neighbourhood N *of* x^0 *such that* $\mathcal{R}_{T|N}$ *is a neighbourhood of* $T(x^0)$, *and there are nonnegative functions* f, *on* $\mathcal{R}_{T|N} \times \Theta$, *and* g, *on* N, *such that*

$$p(x, \theta) = f(T(x), \theta) g(x) \text{ for all } x \in N,\ \theta \in \Theta, \qquad \dots \quad (2.5)$$

and such that $f(y, \theta)$ *has continuous partial derivatives* $\dfrac{\partial f}{\partial y_i}, \dfrac{\partial f}{\partial \theta_j}, \dfrac{\partial^2 f}{\partial y_i \partial \theta_j} = \dfrac{\partial^2 f}{\partial \theta_j \partial y_i}$,

$i = 1, 2, \dots, r$; $j = 1, 2, \dots, \nu$, *in* $\mathcal{R}_{T|N} \times \Theta$ *and* g *has continuous partial derivatives* $\dfrac{\partial g}{\partial x_i}$,

$i = 1, 2, \dots, n$ *in* N.

This lemma plays the crucial part of providing locally continuously differentiable factorizing functions when a locally sufficient statistic is Euclidean and regular. And it implies, moreover, that the factorization holds, for each $\theta \in \Theta$, not merely *almost everywhere* in N, but *everywhere* in N.

Let x be any particular point of Ω. Let $j_1, j_2, \ldots, j_n$ be any particular n integers, each chosen from the set $\{1, 2, \ldots, \nu\}$, and let $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(n)}$ be any particular n points of Θ, not necessarily all distinct. We define the $n \times n$ matrix

$$L(x; j_1, j_2, \ldots, j_n; \theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(n)}) = \left\| \left(\frac{\partial^2 \log p}{\partial x_i \partial \theta_{j_k}}\right)_{x,\, \theta^{(k)}} \right\|_{i,\, k = 1, 2, \ldots, n.} \quad \ldots \quad (2.6)$$

Let $\mathcal{L}_x$ denote the class of all matrices (2.6) for the given point x. We define the integer-valued function ρ_1 on Ω, as follows :

$$\rho_1(x) = \max_{L \in \mathcal{L}_x} (\text{rank } L). \quad \ldots \quad (2.7)$$

And if $\mathcal{S}_{x^0, \alpha}$ denotes the open sphere in Ω of radius α, centered at x^0, we define another integer-valued function ρ on Ω by

$$\rho(x^0) = \lim_{\alpha \downarrow 0} \max_{x \in \mathcal{S}_{x^0, \alpha}} \rho_1(x). \quad \ldots \quad (2.8)$$

We have always

$$\rho_1(x) \leqslant \rho(x), \quad \ldots \quad (2.9)$$

and we make the following definition :

Definition 2.3 : A point x is called a regular point of Ω (for the family $\mathcal{P}$) if equality holds in (2.9).

The set of regular points of Ω is denoted by R. The significant facts about the set R are these :,

Lemma 2.3 : *The set R of regular points of Ω is an everywhere dense, open subset of Ω.*

The function ρ is continuous on R; and, in fact, R is precisely the set of points of continuity of the function ρ_1.

The problem of minimal dimensionality of sufficient statistics has been solved in terms of the function ρ. The solution of the local problem is given by the following two theorems (which are Theorems 3.2 and 3.3 in Barankin and Katz (1959).

Theorem 2.1 : *If T is a statistic of $\mathcal{P}$ which is Euclidean of dimension r at x^0 and continuously differentiable about x^0 (but not necessarily regular at x^0), and if T is sufficient for $\mathcal{P}$ about x^0, then $r \geqslant \rho(x^0)$.*

Theorem 2.2 : *If x^0 is a regular point of Ω, then there exists a sufficient statistic, T, for $\mathcal{P}$ which is Euclidean of dimension $\rho(x^0)$ at x^0, and regular at x^0.*

The proof of the latter theorem gives an explicit construction of such a sufficient statistic T. The following theorem gives the significant aspect of this construction (we give a modified statement of Theorem 3 in Barankin (1960a) or Theorem 2.3 in Barankin (1960b)) :

Theorem 2.3 : *Let x^0 be a regular point of Ω, and let $r_0 = \rho(x^0)$. Let $j_1, j_2, \ldots, j_{r_0}$ be integers, and $\theta^{(1)}, \theta^{(2)}, \ldots \theta^{(r_0)}$ be points of Θ such that the matrix*

$$\left\| \left(\frac{\partial^2 \log p}{\partial x_i \partial \theta_{j_k}} \right)_{x^0, \theta^{(k)}} \right\|_{i = 1, 2, \ldots, n;\; k = 1, 2, \ldots, r_0} \qquad \ldots \quad (2.10)$$

is of rank r_0. Let the functions η_k on Ω accordingly be defined by

$$\eta_k(x) = \left(\frac{\partial \log p(x, \theta)}{\partial \theta_{j_k}} \right)_{\theta^{(k)}}, \qquad k = 1, 2, \ldots, r_0. \qquad \ldots \quad (2.11)$$

Then the statistic T_0, given by

$$T_0(x) = (\eta_1(x), \eta_2(x), \ldots, \eta_{r_0}(x)) \qquad \ldots \quad (2.12)$$

is sufficient for $\mathcal{P}$ about x^0.

Local minimal dimensionality, as presented by Theorems 2.1 and 2.2 above, is achievable in a single globally sufficient statistic simultaneously almost everywhere in R. This is the assertion of Theorem 4.1 in Barankin and Katz (1959), and we restate it here :

Theorem 2.4 : *There exists a sufficient statistic, T^*, for $\mathcal{P}$ which, for almost all (Lebesgue) points $x^0 \in R$, is Euclidean of dimension $\rho\ (x^0)$ at x^0, and regular at x^0,*

The statistic T^* of this theorem is built up (see Barankin and Katz, 1959) from the locally sufficient pieces of the statistics T_0 of Theorem 2.3 above. These pieces, in addition to being dimensionally minimal, are functionally minimal. This fact gave us Theorem 4.3 of Barankin and Katz (1959), which was restated as Theorem 2.5 in Barankin (1960b). But the formulation of the mentioned Theorem 4.3 is somewhat in error, and this was corrected in Barankin (1961). The correction was effected by proving, in fact, the following result :

Theorem 2.5 : *Let j_0 be any particular integer between 1 and ν inclusive, and θ^0 be any particular point of Θ; and define*

$$\eta(x) = \left(\frac{\partial \log p(x, \theta)}{\partial \theta_{j_0}} \right)_{\theta^0} \qquad \ldots \quad (2.13)$$

If T is any sufficient statistic for $\mathcal{P}$, then η is almost everywhere a function of T in Ω.

The correct assertion regarding functional minimality of the statistic T^* is consequently the following (see Section 2 of Barankin, 1961) :

Theorem 2.6 : *The $\mathcal{P}$-sufficient statistic T^* of Theorem 2.4 is locally almost everywhere (Lebesgue) functionally minimal at almost all (Lebesgue) points of R. That is, for each point x in a set $A \subseteq R$, with $R - A$ of measure 0, there is a neighbourhood N of x such that if T is any sufficient statistic then there is a subset C_N of N of measure 0, such that $T^*(x') = T^*(x'')$ whenever x' and x'' are points of $N - C_N$ with $T(x') = T(x'')$.*

3. THE LOCAL RESULTS FOR A FAMILY OF TYPE $\mathscr{P}_{\text{II}}$

Theorem 3.1 below is our basic result relating the existence of a sufficient statistic of dimension $s < n$ to the fact of exponential type of the component densities, for a family of type $\mathscr{P}_{\text{II}}$ defined in (1.2). In anticipation of the proof, we first establish the result of Lemma 3.2 below.

We shall say that there is a *constant linear relation* among finitely many given, real-valued functions on Θ, say $c_1, c_2, \ldots, c_n$, if there exist constants $\alpha_0, \alpha_1, \ldots, \alpha_n$, not all 0, such that

$$\sum_{i=1}^{r} \alpha_i c_i(\theta) \equiv \alpha_0, \quad \theta \in \Theta. \qquad \ldots \quad (3.1)$$

The following lemma has been proved in Barankin (1961), and we quote it here without proof.

Lemma 3.1 : *There is no constant linear relation among the functions* $c_1, c_2, \ldots, c_r$ *on* Θ *if and only if there exist integers* $j_1, j_2, \ldots, j_r$, *each between* 1 *and* ν *inclusive, and corresponding points* $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(r)}$ *of* Θ *such that the matrix*

$$\left\| \left(\frac{\partial c_i}{\partial \theta_{j_k}} \right)_{\theta^{(k)}} \right\|_{i,\, k = 1, 2, \ldots, r} \qquad \ldots \quad (3.2)$$

is nonsingular.

We now have the following lemma.

Lemma 3.2 : *Let* c_i, $i = 1, 2, \ldots, r$ *be real-valued functions on* Θ *and* $\beta_{\lambda i}$, $\lambda = 1, 2, \ldots, r$; $i = 0, 1, 2, \ldots, r$, *be constants; let*

$$b_\lambda(\theta) = \beta_{\lambda 0} + \sum_{i=1}^{r} \beta_{\lambda i} c_i(\theta), \quad \theta \in \Theta \qquad \ldots \quad (3.3)$$

$$= 1, 2, \ldots, r.$$

Then, there is no constant linear relation among the functions b_λ, $\lambda = 1, 2, \ldots, r$, *if and only if the matrix* $\|\beta_{\lambda i}\|_{\lambda,\, i = 1, 2, \ldots, r}$ *is nonsingular and there is no constant linear relation among the functions* c_i, $i = 1, 2, \ldots, r$.

Proof : For any integers $j_1, j_2, \ldots, j_r$, each between 1 and ν inclusive, and any points $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(r)}$ in Θ, we have from (3.3) that

$$\left\| \left(\frac{\partial b_\lambda}{\partial \theta_{j_k}} \right)_{\theta^{(k)}} \right\| = \|\beta_{\lambda i}\| \cdot \left\| \left(\frac{\partial c_i}{\partial \theta_{j_k}} \right)_{\theta^{(k)}} \right\|, \qquad \ldots \quad (3.4)$$

where the indices λ, k, i all range from 1 to r. The product matrix on the right is nonsingular if and only if both factors are nonsingular. It follows that there exist j_k's and $\theta^{(k)}$'s such that the matrix on the left is nonsingular if and only if $\|\beta_{\lambda i}\|$ is nonsingular and there exist j_k's and $\theta^{(k)}$'s such that the second factor on the right is nonsingular. Then, applying Lemma 3.1, we have the asserted result.

We shall say of the expression (1.1) that it is *in reduced form* if there is no constant linear relation among the functions b_λ, $\lambda = 1, 2, \ldots, r$. With this, we are ready to state and prove the basic theorem.

Theorem 3.1 : *Let $\mathscr{P}_{\text{II}}$ be the family of distributions defined in* (1.2), *and let $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$ be a regular point of Ω for the family $\mathscr{P}_{\text{II}}$.*

A necessary and sufficient condition that there exists a sufficient statistic for $\mathscr{P}_{\text{II}}$ about x^0 which is Euclidean of dimension $s < n$ at x^0 and continuously differentiable about x^0 is that for some integer $r \leqslant s$ there be $n-r$ of the factor densities, say $p_{r+1}, p_{r+2}, \ldots, p_n$, with the following three properties :

(i) *they are all of local exponential type as follows :*

$$\log p_m(\xi, \theta) = b_0^{(m)}(\theta)+\psi_0^{(m)}(\xi)+\sum_{\lambda=1}^{r} b_\lambda(\theta)\psi_\lambda{}^{(m)}(\xi), \qquad \ldots \quad (3.5)$$

$$\xi \epsilon \text{ some neighbourhood of } x_m^0;\ \theta \epsilon \Theta;$$
$$m = r+1, r+2, \ldots, n,$$

with all functions appearing in these expressions being continuously differentiable;

(ii) *there is no constant linear relation among the r common parametric functions b_λ, $\lambda = 1, 2, \ldots, r$, that enter into* (3.5); *that is, everyone of the $n-r$ expressions in* (3.5) *is in reduced form;*

(iii) *the functions b_λ, $\lambda = 1, 2, \ldots, r$ are constant linear combinations of the logarithmic derivatives, evaluated at x^0, of the remaining r factor densities, thus :*

$$b_\lambda(\theta) = \beta_{\lambda 0}+ \sum_{i=1}^{r} \beta_{\lambda i}\left(\frac{\partial \log p_i(x_i, \theta)}{\theta x_i}\right)_{x_i=x_i^0} \lambda = 1, 2, \ldots, r, \qquad \ldots \quad (3.6)$$

wherein the matrix $\|\beta_{\lambda i}\|$ is nonsingular.

For $r = 0$ these conditions are to be understood as follows : (3.6) *is vacuous and the r-term sum on the right-hand side of* (3.5) *is* 0.

The integer r for which the above properties are verified is unique and is precisely the local minimal dimension of a continuously differentiable, Euclidean sufficient statistic for $\mathscr{P}_{\text{II}}$ about x^0.

Proof : We shall first establish the last statement of the theorem, namely, that $r = \rho(x^0)$ (see Section 2).

If $r = 0$ we see that each p_i, $i = 1, 2, \ldots, n$, is a product of a function of θ and a function of x_i, for all $\theta \epsilon \Theta$ and all x_i in some neighbourhood of x_i^0. It follows that p itself is a product of a function of θ and a function of x, for all $\theta \epsilon \Theta$ and all x in some neighbourhood of x^0. In this case—see Definition 2.2—a constant function in a neighbourhood of x^0 provides a continuously differentiable, Euclidean sufficient statistic for $\mathscr{P}_{\text{II}}$ about x^0. Thus, $\rho(x^0) = 0 = r$.

Now suppose $r > 0$. Since, by (ii), there is no constant linear relation among the b_λ, $\lambda = 1, 2, \ldots, r$, it follows, on applying Lemma 3.2 to (3.6), both that $\|\beta_{\lambda i}\|$ is nonsingular (—thus, this assertion in (iii) is a consequence of (ii)—) and that the r functions on Θ

$$\left(\frac{\partial \log p_i(x_i, \theta)}{\partial x_i}\right)_{x_i=x_i^0}, \quad i = 1, 2, \ldots, r \qquad \ldots \quad (3.7)$$

have no constant linear relation among them. By Lemma 3.1, then, there exist integers $j_1, j_2, \ldots, j_r$ and points $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(r)}$ in Θ such that the matrix

$$\left\| \left(\frac{\partial^2 \log p_i}{\partial x_i \partial \theta_{j_k}} \right)_{x_i^0,\, \theta^{(k)}} \right\|_{i, k = 1, 2, \ldots, r} \qquad \ldots \quad (3.8)$$

is nonsingular. If we notice that—since p is the product of the p_i and each variable x_i enters into p only through the single factor p_i—we have

$$\frac{\partial \log p}{\partial x_i} \equiv \frac{\partial \log p_i}{\partial x_i}, \quad i = 1, 2, \ldots, n, \qquad \ldots \quad (3.9)$$

then we see that (3.8) is identical with the matrix

$$\left\| \left(\frac{\partial^2 \log p}{\partial x_i \partial \theta_{j_k}} \right)_{x^0,\, \theta^{(k)}} \right\|_{i, k = 1, 2, \ldots, r} \qquad \ldots \quad (3.10)$$

which is therefore nonsingular. Recalling the definition of ρ in (2.8), we conclude that $\rho(x^0) \geqslant r$.

To establish the reverse inequality, let us consider the second mixed partial derivative of log p with respect to x_m and θ_j, for $m \geqslant r+1$. Taking account of (3.9), we obtain this from (3.5). If, moreover, we substitute for the b_λ from (3.6), and again use (3.9), then we have, on evaluating at $x = x^0$,

$$\left(\frac{\partial^2 \log p}{\partial x_m \partial \theta_j} \right)_{x^0} = \sum_{i=1}^{r} \left(\sum_{\lambda=1}^{r} \beta_{\lambda i} \psi_\lambda{}^{(m)\prime}(x_m^0) \right) \left(\frac{\partial^2 \log p}{\partial x_i \partial \theta_j} \right)_{x^0}, \qquad \ldots \quad (3.11)$$

$$m = r+1, r+2, \ldots, n; \; j = 1, 2, \ldots, \nu.$$

From these relations we have the following consequence : for any set of integers $j_1, j_2, \ldots, j_n$ and any points $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(n)}$, the matrix

$$\left\| \left(\frac{\partial^2 \log p}{\partial x_i \partial \theta^{j_k}} \right)_{x^0,\, \theta^{(k)}} \right\|_{i, k = 1, 2, \ldots, n} \qquad \ldots \quad (3.12)$$

is such that each of the last $n-r$ rows is a linear combination of the first r rows. It follows that the rank of this matrix is not greater than r. By the definition of the function ρ_1 (see (2.7)) we have, therefore, that $\rho_1(x^0) \leqslant r$. But x^0 is a regular point, so that $\rho(x^0) = \rho_1(x^0)$. Hence, we have $\rho(x^0) \leqslant r$, and this, combined with the reverse inequality derived above, establishes the equality $\rho(x^0) = r$, as was asserted.

We now turn to the proof of necessity of the conditions (i), (ii) and (iii). Suppose there exists a sufficient statistic for $\mathscr{P}_{\text{II}}$ about the regular point x^0, which is Euclidean of dimensions $s < n$ at x^0 and continuously differentiable about x^0. Then, by Theorem 2.1, we have $\rho(x^0) \leqslant s < n$. For brevity, let us set $r = \rho(x^0)$. Consider first the case of $r = 0$. In this case we have, by virtue of the definition of ρ,

$$\left(\frac{\partial^2 \log p}{\partial x_i \partial \theta_j} \right)_{x^0} = 0, \; \theta \in \Theta, \, i = 1, 2, \ldots, n; \; j = 1, 2, \ldots, \nu. \qquad \ldots \quad (3.13)$$

But, by Lemma 2.3, ρ is identically 0 in a neighbourhood of x^0. Choosing a rectangular neighbourhood, we have, therefore, the following stronger statement :

$$\frac{\partial^2 \log p}{\partial x_i \partial \theta_j} = 0, \quad x \in N_1 \times N_2 \times \dots \times N_n \; ; \; \theta \in \Theta, \qquad \dots \quad (3.14)$$

$$i = 1, 2, \dots, n; \quad j = 1, 2, \dots, \nu,$$

where, for each i, N_i is a 1-dimensional neighbourhood of x_i^0. Bringing (3.9) to bear, the equations (3.14) become

$$\frac{\partial^2 \log p_i(\xi, \theta)}{\partial \xi \partial \theta_j} = 0, \quad \xi \in N_i, \quad \theta \in \Theta, \qquad \dots \quad (3.15)$$

$$i = 1, 2, \dots, n; \qquad j = 1, 2, \dots, \nu.$$

From these it follows that

$$\log\ p_i(\xi, \theta) = b_0^{(i)}(\theta) + \psi_0^{(i)}(\xi), \quad \xi \in N_i,\ \theta \in \Theta,\ i = 1, 2, \dots, n, \qquad \dots \quad (3.16)$$

with the functions $b_0^{(i)}$ and $\psi_0^{(i)}$ being continuously differentiable, by virtue of the differentiability properties of the p_i. This completes the proof of necessity for $r = 0$.

Consider $r > 0$. By a re-indexing of the components $x_1, x_2, \dots, x_n$ if necessary, we have that there exist r points of Θ, say $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(r)}$, and integers $j_1, j_2, \dots, j_r$ such that

$$\left| \left(\frac{\partial^2 \log p}{\partial x_i \partial \theta_{j_k}} \right)_{x^0,\, \theta^{(k)}} \right| \neq 0,\ i, k = 1, 2, \dots, r \qquad \dots \quad (3.17)$$

while

$$\begin{vmatrix} \begin{matrix} \left(\dfrac{\partial^2 \log p}{\partial x_i \partial \theta_{j_k}} \right)_{x^0,\, \theta^{(k)}} \\ i, k = 1, 2, \dots, r \end{matrix} & & \begin{matrix} \left(\dfrac{\partial^2 \log p}{\partial x_1 \partial \theta_j} \right)_{x^0,\, \theta} \\ \vdots \\ \left(\dfrac{\partial^2 \log p}{\partial x_r \partial \theta_j} \right)_{x^0,\, \theta} \end{matrix} \\ \left(\dfrac{\partial^2 \log p}{\partial x_m \partial \theta_{j_1}} \right)_{x^0,\, \theta^{(1)}} \cdots & \left(\dfrac{\partial^2 \log p}{\partial x_m \partial \theta_{j_r}} \right)_{x^0,\, \theta^{(r)}} & \left(\dfrac{\partial^2 \log p}{\partial x_m \partial \theta_j} \right)_{x^0,\, \theta} \end{vmatrix} \equiv 0 \qquad \dots \quad (3.18)$$

for all $\theta \in \Theta$, *every* $j = 1, 2, \dots, \nu$, *and every* $m = r+1, r+2, \dots, n$.

Consider any particular fixed integer m between $r+1$ and n, inclusive. Again by virtue of Lemma 2.3, we have that the determinant (3.18) vanishes identically in θ and j not only at x^0 but for all x in some neighbourhood of x^0. (Since the mixed partial derivatives of p are continuous we can furthermore choose this neighbourhood so small that the determinant (3.17) is nonzero throughout the neighbourhood, not just at x^0.) We may take this neighbourhood to be a direct product of a 1-dimensional neighbourhood, N_m, of x_m^0 and an $(n-1)$-dimensional neighbourhood of $(x_1^0, \dots, x_{m-1}^0, x_{m+1}^0, \dots, x_n^0)$. On

doing this, and restricting attention to only the points $(x_1^0, \ldots, x_{m-1}^0, \xi, x_{m+1}^0, \ldots, x_n^0)$ for $\xi \in N_m$, we get, in place of (3.18), after utilizing (3.9),

$$\begin{vmatrix} \begin{matrix} \left(\dfrac{\partial^2 \log p_i(x_i, \theta')}{\partial x_i \partial \theta'_{j_k}}\right)_{x_i^0,\, \theta^{(k)}} \\ i, k = 1, 2, \ldots, r \end{matrix} & \begin{matrix} \left(\dfrac{\partial^2 \log p_1(x_i \cdot \theta)}{\partial x_1 \partial \theta_j}\right)_{x_1^0} \\ \vdots \\ \left(\dfrac{\partial^2 \log p_r(x_r, \theta)}{\partial x_r \partial \theta_j}\right)_{x_r^0} \end{matrix} \\ \left(\dfrac{\partial^2 \log p_m(\xi, \theta')}{\partial \xi \partial \theta'_{j_1}}\right)_{\theta^{(1)}} \cdots \left(\dfrac{\partial^2 \log p_m(\xi, \theta')}{\partial \xi \partial \theta'_{j_r}}\right)_{\theta^{(r)}} & \dfrac{\partial^2 \log p_m(\xi, \theta)}{\partial \xi \partial \theta_j} \end{vmatrix} \equiv 0 \quad \ldots \quad (3.19)$$

$$\textit{for all } \xi \in N_m, \theta \in \Theta \textit{ and } j = 1, 2, \ldots, \nu.$$

The $r \times r$ subdeterminant in the upper left-hand corner of (3.19) is precisely the determinant (3.17), and is thus nonvanishing. If we expand (3.19) in terms of cofactors of the last row, we obtain equations of the form

$$\bar{b} \frac{\partial^2 \log p_m(\xi, \theta)}{\partial \xi \partial \theta_j} + \sum_{\lambda=1}^{r} \bar{b}_{j\lambda}(\theta) \left(\frac{\partial^2 \log p_m(\xi, \theta')}{\partial \xi \partial \theta'_{j_\lambda}}\right)_{\theta^{(\lambda)}} \equiv 0, \quad \ldots \quad (3.20)$$
$$\xi \in N_m, \quad \theta \in \Theta, \quad j = 1, 2, \ldots, \nu,$$

where $\bar{b}$ and $\bar{b}_{j\lambda}$ are the cofactors in question, and $\bar{b}$ in particular is a nonvanishing constant. On dividing (3.20) by $\bar{b}$ and setting

$$\psi_\lambda^{(m)}(\xi) = \left(\frac{\partial \log p_m(\xi, \theta')}{\partial \theta'_{j_\lambda}}\right)_{\theta^{(\lambda)}}, \; \lambda = 1, 2, \ldots, r; \; m = r+1, r+2, \ldots, n, \quad \ldots \quad (3.21)$$

we may write
$$\frac{\partial^2 \log p_m(\xi, \theta)}{\partial \xi \partial \theta_j} = \frac{\partial}{\partial \xi} \sum_{\lambda=1}^{r} b_{j\lambda}(\theta) \psi_\lambda^{(m)}(\xi), \quad \ldots \quad (3.22)$$
$$\xi \in N_m, \quad \theta \in \Theta, \quad j = 1, 2, \ldots, \nu.$$

Integration of these equations with respect to ξ is immediate, and we get

$$\frac{\partial \log p_m(\xi, \theta)}{\partial \theta_j} = b_{j0}^{(m)}(\theta) + \sum_{\lambda=1}^{r} b_{j\lambda}(\theta) \psi_\lambda^{(m)}(\xi), \quad \xi \in N_m, \; \theta \in \Theta, \quad j = 1, 2, \ldots, \nu. \quad \ldots \quad (3.23)$$

Now, tracing the $b_{j\lambda}$ back to their definition in (3.19), we see that if we define

$$b_\lambda(\theta) = \beta_{\lambda 0} + \frac{(-1)^{r+\lambda}}{\bar{b}} \begin{vmatrix} \begin{matrix} \left(\dfrac{\partial^2 \log p_i(x_i, \theta')}{\partial x_i \partial \theta'_{j_k}}\right)_{x_i^0,\, \theta^{(k)}} \\ i = 1, 2, \ldots, r \\ k = 1, 2, \ldots, \lambda-1, \lambda+1, \ldots, r \end{matrix} & \begin{matrix} \left(\dfrac{\partial \log p_1(x_1, \theta)}{\partial x_1}\right)_{x_1^0} \\ \vdots \\ \left(\dfrac{\partial \log p_r(x_r, \theta)}{\partial x_r}\right)_{x_r^0} \end{matrix} \end{vmatrix}, \quad \ldots \quad (3.24)$$
$$\lambda = 1, 2, \ldots, r,$$

where the $\beta_{\lambda 0}$ are any particular constants, then we have precisely

$$b_{j\lambda}(\theta) = \frac{\partial b_\lambda(\theta)}{\partial \theta_j}, \; j = 1, 2, \ldots, \nu; \; \lambda = 1, 2, \ldots, r. \quad \ldots \quad (3.25)$$

It follows, therefore, that equations (3.23) imply the existence of functions $b_0^{(m)}(\theta)$ and $\psi_0^{(m)}(\xi)$ such that

$$\log p_m(\xi, \theta) = b_0^{(m)}(\theta)+\psi_0^{(m)}(\xi)+\sum_{\lambda=1}^{r} b_\lambda(\theta)\psi_\lambda^{(m)}(\xi),\ \xi \in N_m,\ \theta \in \Theta. \qquad \dots \quad (3.26)$$

The collection of these results for $m = r+1, r+2, \dots, n$ is exactly (3.5), and the b_λ, $\lambda = 1, 2, \dots, r$, here are, as asserted, independent of m. It is clear from (3.21) and (3.24) that the b_λ and the $\psi_0^{(m)}$, $\lambda = 1, 2, \dots, r$, are continuously differentiable. This fact together with the differentiability properties of the function on the left of (3.26) implies that also $b_0^{(m)}$ and $\psi_0^{(m)}$ are continuously differentiable.

It remains to establish assertions (ii) and (iii) of the theorem. To prove (ii), let us take the integers $j_1, j_2, \dots, j_r$ and the points $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(r)}$ that enter into the determinant $\bar{b}$, which is the $r\times r$ subdeterminant in the upper left corner of (3.19). Using (3.24) directly for our evaluations, we readily find the following result :

$$\left(\frac{\partial b_\lambda}{\partial \theta_{j_k}}\right)_{\theta^{(k)}} = \delta_{\lambda k}, \quad \lambda, k = 1, 2, \dots, r, \qquad \dots \quad (3.27)$$

where $\delta_{\lambda k}$ is the Kronecker delta. Thus, for these j_k and $\theta^{(k)}$, the matrix

$$\left\| \left(\frac{\partial b_\lambda}{\partial \theta_{j_k}}\right)_{\theta^{(k)}} \right\| \qquad \dots \quad (3.28)$$

is nonsingular. Therefore, by Lemma 3.1, there is no constant linear relation among the functions $b_1, b_2, \dots, b_r$, and this is what was to be shown.

Finally, to prove (iii), we have only to expand the determinant in (3.24) in terms of cofactors of the last column. And we have already seen above that the nonsingularity of the matrix $\|\beta_{\lambda i}\|$ in (3.6) is a consequence of (ii).

We have now completed our demonstration of the necessity of the conditions (i), (ii) and (iii) in the theorem. We consider finally the proof of sufficiency of these conditions, which can now be given with dispatch.

Suppose conditions (i), (ii) and (iii) are satisfied for some $r \leqslant s$. It has already been shown that then $r = \rho(x^0)$. Therefore, by Theorem 2.2, there is a sufficient statistic for $\mathscr{P}_{\text{II}}$ about x^0 which is Euclidean of dimension r and continuously differentiable about x^0. If $r = s$, the proof is hereby complete. If $r < s$ then we simply adjoin any $s-r$ continuously differentiable, real-valued functions to the sufficient statistic just found, and this provides a local sufficient statistic of dimension s, as asserted.

The proof of Theorem 3.1 is finished.

The point of view taken by Koopman in his theorems is not, as in our theorem above, to consider the question of existence exclusively. It is rather to take a given statistic of dimension s and to ask after conditions that it be a sufficient statistic. Our next theorem is cast in these terms, and is the direct local generalization of

Koopman's results to the case of nonidentical factor densities. (It will be noticed that we assume the statistic T to be locally continuously differentiable, while Koopman assumes merely continuity. What accounts for this is that Koopman's definition of a sufficient statistic is more stringent than the currently accepted definition, an equivalent form of which is given in Lemma 2.1. Thus, he has been able to argue on the weaker hypothesis of continuity of a sufficient statistic, whereas we have required continuous differentiability in order to achieve the fundamental Lemma 2.2, which enables the analysis to proceed from the weaker, modern definition of a sufficient statistic.)

Theorem 3.2 : *Let $\mathcal{P}_{\Pi}$ be the family of distributions defined in* (1.2), *and let* $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$ *be a regular point of Ω for the family $\mathcal{P}_{\Pi}$. Let T be a statistic of $\mathcal{P}_{\Pi}$ which is Euclidean of dimension $s < n$ at x^0 and continuously differentiable about* x^0.

A necessary and sufficient condition that T be sufficient for $\mathcal{P}_{\Pi}$ about x^0 is that for some integer $r \leqslant s$ there be $n-r$ of the factor densities, say $p_{r+1}, p_{r+2}, \ldots, p_n$, for which the properties (i), (ii) *and* (iii) *of Theorem* 3.1 *hold, and furthermore that if $r > 0$ and if $j_1, j_2, \ldots, j_r$ and $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(r)}$ are some particular chosen sets of integers and points of Θ such that*

$$\left| \left(\frac{\partial^2 \log p_i(x_i, \theta)}{\partial x_i \partial \theta j_k} \right)_{x_i^0,\, \theta^{(k)}} \right|_{i, k = 1, 2, \ldots, r} \neq 0; \qquad \ldots \quad (3.29)$$

(*—and these will exist—*), *then the statistics*

$$\sum_{i=1}^{n} \left(\frac{\partial \log p_i(x_i, \theta)}{\partial \theta j_k} \right)_{\theta^{(k)}}, \quad k = 1, 2, \ldots, r \qquad \ldots \quad (3.30)$$

are functions of T almost everywhere (*Lebesgue*) *in some neighbourhood of x^0.*

Of the functions

$$\psi_k^{(i)}(\xi) \equiv \left(\frac{\partial \log p_i(\xi, \theta)}{\partial \theta j_k} \right)_{\theta^{(k)}}, \quad i = 1, 2, \ldots, n; \; k = 1, 2, \ldots, r, \qquad \ldots \quad (3.31)$$

that enter into (3.30), *those with index $i \geqslant r+1$ are precisely the functions $\psi_\lambda^{(m)}$, $\lambda = 1, 2, \ldots, r$; $m = r+1, r+2, \ldots, n$, that figure in one of the possible sets of local representations* (3.5) *of the $p_m(\xi, \theta)$, $m = r+1, r+2, \ldots, n$.*

Proof : We first prove necessity. If T is a sufficient statistic for $\mathcal{P}_{\Pi}$ about x^0, then by Theorem 3.1 there is an integer $r \leqslant s$ such that, with suitable re-indexing of the p_i if necessary, the factor densities $p_{r+1}, p_{r+2}, \ldots, p_n$ have the properties (i), (ii) and (iii) of that theorem. Suppose $r > 0$. According to Theorem 3.1, r is precisely $\rho(x^0)$, and in the necessity proof in that theorem we have seen that we may select integers $j_1, j_2, \ldots, j_r$ and points $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(r)}$ of Θ such that (3.29) holds, and then a set of representations (3.5) is determined, via (3.21) and (3.24), with these j_k and $\theta^{(k)}$. Let some particular such collection of j_k and $\theta^{(k)}$ be chosen. By Theorem 2.3, the functions

$$\eta_k(x) = \left(\frac{\partial \log p(x, \theta)}{\partial \theta j_k} \right)_{\theta^{(k)}}, \qquad k = 1, 2, \ldots, r, \qquad \ldots \quad (3.32)$$

constitute a sufficient statistic for $\mathscr{P}_{\text{II}}$ about x^0, which is Euclidean of minimal dimension r at x^0 and continuously differentiable about x^0. By (3.9) we have

$$\eta_k(x) = \sum_{i=1}^{n} \left(\frac{\partial \log p_i(x_i, \theta)}{\partial \theta_{j_k}} \right)_{\theta^{(k)}}, \quad k = 1, 2, \ldots, r. \qquad \ldots \quad (3.33)$$

Theorem 2.5 asserts that $(\eta_1, \eta_2, \ldots, \eta_r)$ is locally almost everywhere (Lebesgue) functionally minimal at x^0. Hence, it is a function of T almost everywhere in some neighbourhood of x^0. This statement is, as we see through (3.33), the asserted result that the functions (3.30) are functions of T almost everywhere about x^0.

Finally, (3.21) in the necessity proof of Theorem 3.1 immediately establishes the last statement of our present theorem. And with this the proof of necessity is complete.

We must now prove the sufficiency of the conditions of our theorem. Having (i), (ii) and (iii) of Theorem 3.1 satisfied, we have, by that theorem that $r = \rho(x^0)$, and furthermore, from the details in the proof of this fact we know that there exist $j_1, j_2, \ldots, j_r$ and $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(r)}$ such that (3.29) holds. It then follows that the η_k in (3.33), which is to say the functions (3.30), constitute a sufficient statistic for $\mathscr{P}_{\text{II}}$ about x^0. It is then a consequence of Definition 2.2 that since these η_k are functions of T almost everywhere in some neighbourhood of x^0, T is itself a sufficient statistic for $\mathscr{P}_{\text{II}}$ about x^0.

We have completed the proof of Theorem 3.2.

We may now give two corollaries, one of each of the above two theorems, which specialize the results to the classical case of identical factor densities.

Corollary 3.1 : *Let $\mathscr{P}_{\text{II}I}$ be the family of distributions defined in* (1.2), *wherein, in particular, the p_i, $i = 1, 2, \ldots, n$, are identically the same function, say p_0. Let Ω_0 be a standard copy of the identical open sets $\Omega_1, \Omega_2, \ldots, \Omega_n$ (see the Introduction), so that $p_0(\xi, \theta)$ is defined on $\Omega_0 \times \Theta$. Let $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$ be a regular point of Ω for the family $\mathscr{P}_{\text{II}I}$.*

A necessary and sufficient condition that there exists a sufficient statistic for $\mathscr{P}_{\text{II}I}$ about x^0 which is Euclidean of dimension $s < n$ at x^0 and continuously differentiable about x^0 is that for some integer $r \leqslant s$ there be $n-r$ of the components of x^0, say $x_{r+1}^0, x_{r+2}^0, \ldots, x_n^0$ with the following properties :

(i) *there is an open set $A \subseteq \Omega_0$ which contains the points $x_{r+1}^0, x_{r+2}^0, \ldots, x_n^0$ and such that p_0 is of exponential type in A :*

$$\log\, p_0(\xi, \theta) = b_0(\theta) + \psi_0(\xi) + \sum_{\lambda=1}^{r} b_\lambda(\theta)\psi_\lambda(\xi), \quad \xi \in A, \quad \theta \in \Theta, \qquad \ldots \quad (3.34)$$

with all functions appearing in this expression being continuously differentiable ;

(ii) *there is no constant linear relation among the r functions b_λ, $\lambda = 1, 2, \ldots, r$, in* (3.34); *that is, the expression on the right in* (3.34) *is in reduced form;*

(iii) *the functions* b_λ, $\lambda = 1, 2, \ldots, r$ *are constant linear combinations of the logarithmic derivative of* p_0 *evaluated at each of the remaining* r *coordinate points,* $x_1^0, x_2^0, \ldots, x_r^0$, *of the* x^0, *thus* :

$$b_\lambda(\theta) = \beta_{\lambda 0} + \sum_{i=1}^{r} \beta_{\lambda i} \left(\frac{\partial \log p_0(\xi, \theta)}{\partial \xi} \right)_{\xi = x_i^0}, \quad \lambda = 1, 2, \ldots, r, \qquad \ldots \quad (3.35)$$

where the matrix $\|\beta_{\lambda i}\|$ *is nonsingular.*

For $r = 0$ *these conditions are to be understood as follows* : (3.35) *is vacuous and the r-term sum on the right-hand side of* (3.34) *is* 0.

The integer r *for which the above properties are verified is unique and is precisely the local minimal dimension of a continuously differentiable, Euclidean sufficient statistic for* $\mathscr{P}_{\Pi I}$ *about* x^0.

Proof : It is necessary only to prove that the three conditions stated here above are equivalent to the three conditions in the statement of Theorem 3.1 when the p_i are all identically equal to p_0. That the present conditions imply those of Theorem 3.1 is immediate; in fact, for each $m = r+1, r+2, \ldots, n$, the set A itself may be taken to be the neighbourhood of x_m^0 for which (3.5) is to hold, and the $n-r$ equalities (3.5) may be taken all identical with (3.34). As for the remaining conditions, the correspondence is clear.

Suppose, conversely, that we have the conditions of Theorem 3.1 verified, with every p_i equal to p_0. Then, clearly, the statements there about the $n-r$ factor densities $p_{r+1}, p_{r+2}, \ldots, p_n$ translate immediately into statements concerning the single density p_0 and its behaviour about the $n-r$ points $x_{r+1}^0, x_{r+2}^0, \ldots, x_n^0$ of Ω_0. We must show that the several local forms (3.5), valid for respective neighbourhoods N_m, say, of the x_m^0, $m = r+1, r+2, \ldots, n$, can be so designated that they constitute a single statement of the form (3.34). This will be so if the designations can be made so that, for each $\lambda = 0, 1, \ldots, r$, the two functions $\psi_\lambda^{(m)}$ and $\psi_\lambda^{(m')}$ agree on $N_m \cap N_{m'}$, for every pair m, m', and so that $b_0^{(m)}$ is independent of m. For, clearly, if these properties are secured then (3.5) goes over into the single statement (3.34) with $A = \bigcup_{m=r+1}^{n} N_m$. Now, in the proof of Theorem 3.1 we have seen that the $\psi_\lambda^{(m)}$, for $\lambda = 1, 2, \ldots, r$ and all $m = r+1, r+2, \ldots, n$, may be taken as defined by (3.21). In our present situation of $p_i \equiv p_0$ for all i, this definition entails immediately that these $\psi_\lambda^{(m)}$ which are defined over all of Ω_0, do not depend on m; hence, they will indeed agree on intersections $N_m \cap N_{m'}$. Furthermore, on such an intersection the two pertinent equalities of (3.5), namely,

$$\log p_0(\xi, \theta) = b_0^{(m)}(\theta) + \psi_0^{(m)}(\xi) + \sum_{\lambda=1}^{r} b_\lambda(\theta) \psi_\lambda(\xi),$$

$$\ldots \quad (3.36)$$

$$\log p_0(\xi, \theta) = b_0^{(m')}(\theta) + \psi_0^{(m')}(\xi) + \sum_{\lambda=1}^{r} b_\lambda(\theta) \psi_\lambda(\xi),$$

then yield $\quad b_0^{(m)}(\theta) + \psi_0^{(m)}(\xi) = b_0^{(m')}(\theta) + \psi_0^{(m')}(\xi), \ \xi \in N_m \cap N_{m'}, \ \theta \in \Theta. \qquad \ldots \quad (3.37)$

From these relations for all pairs m and m' it follows that the functions $b_0^{(m)}$ *on* θ differ by at most additive constants, and likewise the functions $\psi_0^{(m)}$ on intersections of their domains. Specifically, there are constants $a_{r+1}, a_{r+2}, \ldots, a_n$ such that $b_0^{(m)}(\theta) \equiv b_0^{(n)}(\theta)+a_m$, $m = r+1, r+2, \ldots, n$; and these constants are furthermore such that the functions $\psi_0^{(m)}(\xi)+a_m$ agree on intersections of their domains. Hence, if, for each m, the functions $b_0^{(m)}$ and $\psi_0^{(m)}$ are replaced by $b_0^{(m)}-a_m$ and $\psi_0^{(m)}+a_m$, respectively, then we have designations of the first two terms on the right-hand side of (3.5) such that the first does not depend on $m(b_0^{(m)}-a_m \equiv b_0^{(n)})$, and the second terms, for different m, agree on intersections of their domains. This achieved, it now follows, as indicated above, that the equalities (3.5) combine into the single statement (3.34) with $A = \bigcup_{m=r+1}^{n} N_m$, by defining $b_0 = b_0^{(n)}$ and $\psi_0 = \psi_0^{(m)}+a_m$ on N_m, $m = r+1, r+2, \ldots, n$, and $\psi_\lambda = \psi_\lambda^{(m)}$ for all $m \geqslant r+1$ and all $\lambda > 0$.

The proof of the corollary is complete.

Corollary 3.2 : *Let $\mathcal{P}_{\Pi I}$ be the family of distributions defined in* (1.2), *wherein, in particular, the p_i, $i = 1, 2, \ldots, n$, are identically the same function, say p_0. Let Ω_0 be as in the statement of Corollary* 3.1, *and let $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$ be a regular point of Ω for $\mathcal{P}_{\Pi I}$. Let T be a statistic of $\mathcal{P}_{\Pi I}$ which is Euclidean of dimension $s < n$ at x^0 and continuously different about x^0.*

A necessary and sufficient condition that T be sufficient for $\mathcal{P}_{\Pi I}$ about x^0 is that for some integer $r \leqslant s$ there be $n-r$ of the components of x^0, say $x_{r+1}^0, x_{r+2}^0, \ldots, x_n^0$, for which the properties (i), (ii) *and* (iii) *of Corollary* 3.1 *hold, and furthermore that, if $r > 0$ and if $j_1, j_2, \ldots, j_r$ and $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(r)}$ are some particular chosen set of integers and points of Θ such that*

$$\left| \left(\frac{\partial^2 \log p_0(\xi, \theta)}{\partial \xi \partial \theta_{j_k}} \right)_{x_i^0,\ \theta^{(k)}} \right|_{i, k = 1, 2, \ldots, r} \neq 0 ; \qquad \ldots \quad (3.38)$$

(–and these will exist–), then the statistics

$$\sum_{i=1}^{n} \left(\frac{\partial \log p_0(\xi, \theta)}{\partial \theta_{j_k}} \right)_{x_i,\ \theta^{(k)}}, \quad k = 1, 2, \ldots, r, \qquad \ldots \quad (3.39)$$

are functions of T almost everywhere (Lebesgue) in some neighbourhood of x^0.

The functions

$$\psi_k(\xi) \equiv \left(\frac{\partial \log p_0(\xi, \theta)}{\partial \theta_{j_k}} \right)_{\theta^{(k)}}, \quad k = 1, 2, \ldots, r \qquad \ldots \quad (3.40)$$

that enter into (3.39) *are precisely the functions ψ_λ, $\lambda = 1, 2, \ldots, r$, that figure in one of the possible representations* (3.34) *of p_0 for some open subset A of Ω_0 containing the points $x_{r+1}^0, x_{r+2}^0, \ldots, x_n^0$.*

Proof : This corollary obtains immediately from Theorem 3.2 by specializing the statements of that theorem to $\mathcal{P}_{\Pi I}$.

It is worth pointing out explicitly that the more complicated statement above involving (3.39) and (3.40) cannot be replaced, as might first be suspected, by the simpler statement that the statistics

$$\sum_{i=1}^{n} \psi_\lambda(x_i), \quad \lambda = 1, 2, \ldots, r, \qquad \ldots \quad (3.41)$$

where the ψ_λ *here are those appearing in* (3.34), are functions of T almost everywhere in some neighbourhood of x^0. The reason for this is that while the functions defined by (3.40) over all of Ω_0 will provide a representation (3.34) for a subset A of Ω_0, the functions ψ_λ as given by (3.34) are defined only over A, and even if they are defined over all of Ω_0, their definition over $\Omega_0 - A$ may be completely irrelevant to present questions, so that the first r terms in each of the sums (3.41) may be arbitrary and meaningless.

4. Analytic density functions

Preparatory to obtaining, in Section 5, what global results can be stated at our intended level of generality in the case of analytic product densities, we now withdraw for a moment from our concentration on product densities to consider analytic density functions generally, be they of the product type or not. Thus, we consider a family $\mathcal{P} = \{\mu_\theta, \theta \in \Theta\}$ of probability measures on the Lebesgue subsets of an open set Ω in E_0^n, a Euclidean n-space. The index set Θ is an open set in a Euclidean ν-space, E_0^ν. Each μ_θ is absolutely continuous with respect to Lebesgue measure, and we suppose there is a determination, p, of the family density function such that $p(x, \theta) > 0$ throughout $\Omega \times \Theta$, and such that for a fixed open, connected set $\Omega' \subseteq \Omega$, p is analytic in $\Omega' \times \Theta$. The set Ω' might, in particular, be all of Ω; but we need not insist on this for our present deliberations.

We shall avail ourselves of the following standard property of analytic functions:

Lemma 4.1 : *If two real-valued functions are analytic in an open connected subset* Υ *of a Euclidean space, and they are equal at all points of a (nonempty) open subset of* Υ, *then they are equal everywhere in* Υ.

The statement and proof of this fact may be found on p. 202 of Dieudonné (1960).

Consider a regular point, x^0, in Ω', and for brevity set $\rho(x^0) = r_0$. Suppose first that $r_0 > 0$. Then, with a re-indexing of the x_i if necessary, there exist integers $j_1, j_2, \ldots, j_{r_0}$ between 1 and ν inclusive, and points $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(r_0)}$ of Θ such that the matrix

$$\Delta(x) = \left\| \left(\frac{\partial^2 \log p}{\partial x_i \, \partial \theta_{j_k}} \right)_{x,\, \theta^{(k)}} \right\| \; i, k = 1, 2, \ldots, r_0 \qquad \ldots \quad (4.1)$$

is nonsingular at x^0. Now, det $\Delta(x)$ is analytic in Ω'. It can, therefore, not vanish in an open subset of Ω'; for, if it did, then by Lemma 4.1 it would vanish everywhere in Ω', and in particular it would vanish at x^0, contradicting the established fact that $\Delta(x^0)$ is nonsingular. Thus, the points at which Δ is nonsingular are dense in Ω', and it

follows, by the definition of the function ρ, that $\rho(x) \geqslant r_0$ for all $x \in \Omega'$. If we now consider that our choice of x^0 is such that r_0 is the largest value that ρ takes on at a regular point in Ω', and if we take account of the fact that the value of ρ at a nonregular point does not exceed the value of ρ at a suitable, arbitrarily nearby regular point, then we obtain our ultimate conclusion : the function ρ is constant on Ω'. If $r_0 = 0$, then the assumption that $\rho(x^1) > 0$ for some regular point x^1 in Ω' leads, by the above argument, to the conclusion that ρ has a constant positive value everywhere in Ω', thus contradicting the fact that $\rho(x^0) = r_0 = 0$. Hence ρ vanishes at all regular points of Ω', and so it vanishes everywhere in Ω'. The conclusion that ρ is constant on Ω' therefore holds in all cases. Thus, the local least possible dimension of a continuously differentiable, Euclidean $\mathscr{P}$-sufficient statistic is the same at all points of Ω' (see Theorems 2.1 and 2.2).

Let us continue to denote by r_0 the constant value of ρ on Ω'. Again suppose first that $r_0 > 0$. Let x^0 be any particular regular point of Ω'. Consider the matrix $\Delta(x)$ of (4.1) constructed with reference to the point x^0; that is, so that $\Delta(x^0)$ is nonsingular. Then, by Theorem 2.3, the r_0 functions

$$\eta_k(x) = \left(\frac{\partial \log p(x, \theta)}{\partial \theta_{j_k}}\right)_{\theta^{(k)}}, \quad k = 1, 2, \ldots, r_0 \qquad \ldots \quad (4.2)$$

constitute a sufficient statistic for $\mathscr{P}$ about x^0. Now, in the argumentation above we have seen that there is a dense (open) set of regular points in Ω' at each of which Δ is nonsingular. Hence, the same set of functions (4.2) constitute (among all continuously differentiable, Euclidean statistics) a minimal dimensional, Euclidean sufficient statistic about each of a dense subset of regular points of Ω'.

If $r_0 = 0$, then a constant function supplies such a statistic, and it is, in fact, of minimal dimension at every regular point of Ω'. Moreover, in this case every point of Ω' is a regular point. This is a ready consequence of the nature of the function ρ.

Let $T_0 = (\eta_1, \eta_2, \ldots, \eta_{r_0})$, supposing that $r_0 > 0$. The question presents itself immediately whether or not T_0 is in fact sufficient for $\mathscr{P}$ in Ω'. According to our conclusions thus far, if x^1 is any particular one of a certain dense subset of regular points of Ω', there exist functions f^1 and g^1 (with restricted domains of definition) such that the relation

$$p(x, \theta) = f^1(T_0(x), \theta) g^1(x) \qquad \ldots \quad (4.3)$$

is valid for all $\theta \in \Theta$ and all x *in some neighbourhood* N^1 *of* x^1; and additionally by virtue of Lemma 2.2 the functions f^1 and g^1 may be taken to have differentiability properties. This relation establishes T_0 as a sufficient statistic for $\mathscr{P}$ about x^1. The question that has been raised is whether or not in fact there exist functions f and g, defined for all Ω' and Θ, so that (4.3)—with f and g in place of f^1 and g^1, respectively—holds for all $\theta \in \Theta$ and *everywhere* in Ω' or at least almost everywhere (Lebesgue) in Ω' for each $\theta \in \Theta$ (see Definition 2.1). It appears that the answer to this question is not always in the affirmative. We must content ourselves with asserting less than this in the general case.

Consider (4.3) in particular for the regular point x^0 which led us to the functions (4.2)—that is, to the vector function T_0. For this point let f^0 and g^0 be the

pertinent functions in (4.3), and let N^0 be the pertinent neighbourhood of x^0. Then we have, about x^0, for all $\theta \epsilon \Theta$,

$$\frac{\partial \log p(x, \theta)}{\partial \theta_j} = \frac{\partial \log f^0(T_0(x), \theta)}{\partial \theta_j}, \quad j = 1, 2, \ldots, \nu. \qquad \ldots \quad (4.4)$$

From these equations we conclude that the partial derivatives $\frac{\partial \log f^0(T_0(x), \theta)}{\partial \theta_j}$ are analytic functions on $N^0 \times \Theta$, and it follows that $f^0(T_0(x), \theta)$ may be taken analytic in $N^0 \times \Theta$. Then, from the relation

$$\log p(x, \theta) = \log f^0(T_0(x), \theta) + \log g^0(x), \qquad \ldots \quad (4.5)$$

it follows that g^0 is analytic in N^0.

Now suppose f^0 has a nonvanishing extension f over $\mathcal{R}_{T_0|\Omega'} \times \Theta'$—where $\mathcal{R}_{T_0|\Omega'}$ denotes the range T_0 restricted to Ω'—such that $f(T_0(x), \theta)$ is analytic in $\Omega' \times \Theta$. Then the analytic derivatives $\frac{\partial \log f(T_0(x), \theta)}{\partial \theta_j}$ are identical with the right-hand sides of (4.4) in $N^0 \times \Theta$, and are consequently equal to the left-hand sides of (4.4) in $N^0 \times \Theta$. Since the left-hand sides of (4.4) and the derivatives $\frac{\partial \log f(T_0(x), \theta)}{\partial \theta_j}$ are all analytic in $\Omega' \times \Theta$, it follows by Lemma 4.1, that

$$\frac{\partial \log f(T_0(x), \theta)}{\partial \theta_j} = \frac{\partial \log p(x, \theta),}{\partial \theta_j}, j = 1, 2, \ldots, \nu; (x, \theta) \epsilon\, \Omega' \times \Theta. \qquad \ldots \quad (4.6)$$

Note that Θ need not be connected, and therefore $\Omega' \times \Theta$ may not be connected. But in this case (4.6) is established by reasoning separately with each connected component of Θ. From (4.6) it follows that

$$p(x, \theta) = f(T_0(x), \theta) g(x), \quad (x, \theta) \epsilon\, \Omega' \times \Theta \qquad \ldots \quad (4.7)$$

for some function g, and as reasoned above, the function g is analytic in Ω'.

Thus, we have shown that if f^0 has a nonvanishing extension f over $\mathcal{R}_{T_0|\Omega'} \times \Theta$ such that $f(T_0(x), \theta)$ is analytic in $\Omega' \times \Theta$, then T_0 is sufficient for $\mathcal{P}$ in Ω', and moreover the factorization (4.7) is in terms of analytic functions.

In the case $r_0 = 0$, the function T_0 as given by (4.2) is to be replaced by $T_0 =$ a fixed numerical constant. And in this case the argumentation proceeding from (4.4) is still valid, so that the statistic $T_0 =$ a fixed numerical constant is sufficient in Ω' if, for some analytic factorization (4.13), f^0 has a nonvanishing extension f over $\mathcal{R}_{T_0|\Omega'} \times \Theta$ such that $f(T_0(x), \theta)$ is analytic over $\Omega' \times \Theta$. But in this case there is such an extension immediately evident : since T_0 is constant, with value a, say, defined on $N^0 \times \Theta$, $f^0(T_0(x), \theta)$ is an analytic function of θ alone, throughout Θ; therefore, define $f(T_0(x), \theta) = f^0(a, \theta)$ for all $(x, \theta) \epsilon\, \Omega' \times \Theta$. The function f is the desired extension. And so, in the case of $r_0 = 0$, a statistic of the form $T_0 =$ a fixed numerical constant is sufficient for $\mathcal{P}$ in Ω'.

We now sum up the results of this section in the following theorem.

Theorem 4.1 : *Let p be a family density function as described at the beginning of this section and, as hypothesized, let Ω' be an open connected subset of Ω such that p is analytic in $\Omega' \times \Theta$. Then the following facts hold :*

(i) *The function ρ is constant over Ω'.*

(ii) *If r_0 is the constant value of ρ in Ω', and if x^0 is any particular regular point of Ω', and if $T_0 = (\eta_1, \eta_2, \ldots, \eta_{r_0})$ is the analytic (over Ω') function given by* (4.2) *in case $r_0 > 0$, and $T_0 =$ a fixed numerical constant in case $r_0 = 0$, then T_0 is, among all continuously differentiable, Euclidean statistics, a minimal dimensional, Euclidean sufficient statistic for $\mathcal{P}$ about each point of a dense subset R_0' of regular points of Ω'. In particular, $R_0' = \Omega'$ in case $r_0 = 0$.*

(iii) *p has an analytic factorization in terms of T_0 about x^0 (and equally well, about each point of R_0'). That is, for some neighbourhood N^0 of x^0,*

$$p(x, \theta) = f^0(T(x), \theta)g^0(x), \quad (x, \theta) \in N^0 \times \Theta, \qquad \ldots \quad (4.8)$$

where $f^0(T(x), \theta)$ is analytic over $N^0 \times \Theta$ and g is analytic over N^0.

(iv) *T_0 is a sufficient statistic for $\mathcal{P}$ in Ω' if, for some analytic factorization* (4.8), *the function f^0 on $\mathcal{R}_{T_0|N^0} \times \Theta$ has a nonvanishing extension f over $\mathcal{R}_{T_0|\Omega} \times \Theta$ such that $f(T_0(x), \theta)$ is analytic over $\Omega' \times \Theta$. And then we have*

$$p(x, \theta) = f(T_0(x), \theta)g(x), \ (x, \theta) \in \Omega' \times \Theta \qquad \ldots \quad (4.9)$$

with g also analytic over Ω'. In particular, in the case $r_0 = 0$ there is such an extension, so that a statistic of the form $T_0 =$ a fixed numerical constant is, in this case, sufficient for $\mathcal{P}$ in Ω'.

5. The global results for a (partially or fully) analytic family of type $\mathcal{P}_{\Pi}$

We are now ready to combine the local theorems of Section 3 with the general consequences of analyticity as seen in Section 4, to obtain the desired global results. We consider a product family $\mathcal{P}_{\Pi}$ as given by (1.2), and such that, furthermore, for each $i = 1, 2, \ldots, n$ there is an open, connected subset (i.e., an open interval) Ω_i' of Ω_i such that p_i is analytic in $\Omega_i' \times \Theta$. Then $\Omega' = \Omega_1' \times \Omega_2' \times \ldots \times \Omega_n'$ is an open, connected subset of Ω, and p is analytic in $\Omega' \times \Theta$. It may be that each Ω_i is itself an interval and that p_i is analytic in $\Omega_i \times \Theta$ for each i, in this case we may take $\Omega_i' = \Omega_i$ for each i, and therefore $\Omega' = \Omega$. Or, it may be that some or all of the Ω_i are not connected, but p_i is analytic in Ω_i for each i, our results below can be applied in this case through separate consideration of the several connected components of the Ω_i. Our investigations in the present section will be concerned exclusively with the subset $\Omega' \times \Theta$ of $\Omega \times \Theta$; and outside this subset p may or may not be analytic. That is, in general, we are directing our attention to a part of the domain of p, which may be a proper subset, on which p is analytic while it may not be analytic in the complementary part of the domain. Thus, the family $\mathcal{P}_{\Pi}$ may be only partially analytic, and the pertinence of our term "global" is to just the analytic part of the domain. On the other hand, $\mathcal{P}_{\Pi}$ may be fully analytic—that is, p may be analytic over all of $\Omega \times \Theta$—and in this case our stated results can be applied, as we have just remarked, for each connected component of Ω separately and "global" then can be warrantedly understood to refer to the full domain of p.

The result nearest at hand is the following :

Theorem 5.1 : *Let* $\mathscr{P}_{\Pi}$ *be a product family of distributions as defined by* (1.2), *with the additional property that* p *is analytic in* $\Omega' \times \Theta$ *where* $\Omega' = \Omega'_1 \times \Omega'_2 \times \ldots \times \Omega'_n$ *is an* n*-dimensional open interval and the* $\Omega'_i \times \Theta$ *are sets of analyticity of the respective factors of* p, *as described above.*

A necessary condition that there exists a statistic which is sufficient for $\mathscr{P}_{\Pi}$ *in* Ω' *and is Euclidean of dimension* $s < n$ *in* Ω', *and which is continuously differentiable about some regular point of* Ω', *is that for some integer* $r \leqslant s$ *there be* $n-r$ *of the factor densities, say* $p_{r+1}, p_{r+2}, \ldots, p_n$, *with the following three properties :*

(i) *the* p_m, $m = r+1, r+2, \ldots, n$, *are of exponential type in the respective domains* $\Omega'_m \times \Theta$, *as follows :*

$$\log p_m(\xi, \theta) = b_0^{(m)}(\theta) + \psi_0^{(m)}(\xi) + \sum_{\lambda=1}^{r} b_\lambda(\theta)\psi_\lambda^{(m)}(\xi), \qquad \ldots \quad (5.1)$$

$$(\xi, \theta) \in \Omega'_m \times \Theta; \quad m = r+1, r+2, \ldots, n,$$

with all functions appearing in these expressions being analytic in the pertinent domain $\Omega'_m \times \Theta$ (*or, equivalently, in* Ω'_m *or in* Θ, *respectively*);

(ii) *there is no constant linear relation among the* r *common parametric functions* b_λ, $\lambda = 1, 2, \ldots, r$, *that enter into the* $n-r$ *expressions* (5.1); *that is, every one of these expressions is in reduced form* ;

(iii) *for suitable representations* (5.1) *of the* p_m, $m = r+1, r+2, \ldots, n$, *the functions* b_λ, $\lambda = 1, 2, \ldots, r$ *are constant linear combinations of the logarithmic derivatives, evaluated at some regular point* $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$ *of the remaining* r *factor densities, thus :*

$$b_\lambda(\theta) = \beta_{\lambda 0} + \sum_{i=1}^{r} \beta_{\lambda i} \left(\frac{\partial \log p_i(x_i, \theta)}{\partial x_i} \right)_{x_i = x_i^0}, \quad \lambda = 1, 2, \ldots, r, \qquad \ldots \quad (5.2)$$

wherein the matrix $\|\beta_{\lambda i}\|$ *is nonsingular.*

For $r = 0$ *these conditions are to be understood as follows* : (5.2) *is vacuous and the* r*-term sum on the right-hand side of* (5.1) *is* 0.

The integer r *for which the above properties are verified is unique and is precisely the constant value of the function* ρ *in* Ω'.

Proof : Suppose there is a statistic which is sufficient for $\mathscr{P}_{\Pi}$ in Ω' and is Euclidean of dimension $s < n$ in Ω', and which is continuously differentiable about some regular point of Ω', say $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$. Then, *a fortiori*, this statistic is sufficient for $\mathscr{P}_{\Pi}$ about x^0. Consequently, Theorem 3.1 applies and (with suitable re-indexing if necessary) the p_m, $m = r+1, r+2, \ldots, n$, have the local exponential form (3.5) described by (i), (ii) and (iii) of that theorem. By virtue of the present analyticity assumptions we see by (3.21) and (3.24) that the functions b_λ and $\psi_\lambda^{(m)}$ in (3.5) are analytic in, respectively, Θ and the pertinent Ω'_m. It follows that also the $b_0^{(m)}$ and $\psi_0^{(m)}$ are analytic in Θ and Ω'_m, respectively. Thus, the equations (3.5), considered in turn for each open, connected component of Θ, state the equality, in an open subset, of two analytic functions over an open, connected set. Therefore, by Lemma 4.1, the equality holds over the entire open, connected set. Aggregating these

results for all the components of Θ, we obtain the assertion (i) of the present theorem, giving the exponential form of p_m over all of $\Omega'_m \times \Theta$, for each $m = r+1, r+2, \ldots, n$.

Assertions (ii) and (iii) of the present theorem are immediate consequences of (ii) and (iii) of Theorem 3.1.

Consider finally the last assertion of the present theorem. Conditions (i), (ii) and (iii), applied locally about the regular point x^0 in question, are precisely the conditions (i), (ii) and (iii) of Theorem 3.1. Therefore, by the last statement of that theorem, we have necessarily $r = \rho(x^0)$. But by Theorem 4.1, ρ is constant over Ω'. Hence, r is, as asserted, the constant value of ρ over Ω'.

This completes the proof of Theorem 5.1.

The next two theorems give circumstances under which the three conditions of Theorem 5.1 are sufficient as well as necessary.

Theorem 5.2 : *Let $\mathcal{P}_{\Pi}$ be a product family of distributions as described in the hypothesis of Theorem* 5.1. *Let conditions* (i), (ii) *and* (iii) *of that theorem hold for some $r \leqslant s < n$ and for some regular point $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$. If $r = 0$, a statistic of the form $T_0 =$ a fixed numerical constant is sufficient for $\mathcal{P}_{\Pi}$ in Ω', and it is analytic and of dimension $r = 0$ over Ω'. Therefore, any statistic which is Euclidean of dimension $s \geqslant r = 0$ in Ω' and is continuously differentiable about some regular point of Ω' is sufficient for $\mathcal{P}_{\Pi}$ in Ω'.*

If $r > 0$, then for suitable integers $j_1, j_2, \ldots, j_r$ and points $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(r)}$ of Θ, the statistic $T_0 = (\eta_1, \eta_2, \ldots, \eta_r)$, where

$$\eta_k(x) = \sum_{i=1}^{n} \left(\frac{\partial \log p_i(x_i, \theta)}{\partial \theta_{jk}} \right)_{\theta^{(k)}} \quad k = 1, 2, \ldots, r, \qquad \ldots \quad (5.3)$$

is sufficient for $\mathcal{P}_{\Pi}$ in some neighbourhood N^0 of x^0 and there are functions f^0 and g^0 such that

$$p(x, \theta) = f^0(T_0(x), \theta) g^0(x), \quad (x, \theta) \in N^0 \times \Theta \qquad \ldots \quad (5.4)$$

and such that $f^0(T_0(x), \theta)$ and $g^0(x)$ are analytic over $N^0 \times \Theta$. The statistic T_0 is analytic over Ω'; and of the functions (which are all analytic over the respective Ω_i''s)

$$\psi_k^{(i)}(\xi) \equiv \left(\frac{\partial \log p_i(\xi, \theta)}{\partial \theta_{jk}} \right)_{\theta^{(k)}} \quad i = 1, 2, \ldots, n; \quad k = 1, 2, \ldots, r, \qquad \ldots \quad (5.5)$$

that enter into T_0, those with index $i \geqslant r+1$ are precisely the functions $\psi_\lambda^{(m)}$, $\lambda = 1, 2, \ldots r$; $m = r+1, r+2, \ldots, n$, that figure in one of the possible sets of representations (5.1) *of the p_m, $m = r+1, r+2, \ldots, n$.*

Now, if the function f^0, defined on $\mathcal{R}_{T_0|N^0} \times \Theta$, has a nonvanishing extension f over $\mathcal{R}_{T_0|\Omega'} \times \Theta$ such that $f(T_0(x), \theta)$ is analytic over $\Omega' \times \Theta$, then T_0—analytic and of dimension r over Ω'—is a sufficient statistic for $\mathcal{P}_{\Pi}$ in Ω'. And it follows, a fortiori, that there is a statistic which is sufficient for $\mathcal{P}_{\Pi}$ in Ω', is Euclidean of dimension $s \geqslant r$ in Ω', and is continuously differentiable about some regular point of Ω'.

Proof : Under the given hypothesis, if $r = 0$, then the very last statement of Theorem 4.1 gives us that $T_0 =$ a fixed numerical constant is a sufficient statistic for $\mathcal{P}_{\Pi}$ in Ω'. Since any statistic of which a sufficient statistic is (almost everywhere) a function is likewise sufficient, and since $T_0 =$ a constant is a function of any function

on Ω', it follows in particular, as asserted, that any statistic which is Euclidean of dimension $s \geqslant r = 0$ in Ω' and is continuously differentiable about some regular point of Ω' is sufficient for $\mathscr{P}_{\text{II}}$ in Ω'.

If $r > 0$, then (5.3) and (5.4), and their attendant remarks, follow from Theorem 4.1. And the assertion regarding (5.5) is a consequence of Theorem 3.2. That T_0 is sufficient for $\mathscr{P}_{\text{II}}$ in Ω' if f^0 has the indicated extension f is an immediate consequence of (iv) of Theorem 4.1. And finally, the very last statement of the present theorem is established by augmenting T_0 with any $s-r$ continuously differentiable components (if $s < r$).

This completes the proof of Theorem 5.2.

Next, we have sufficient conditions as follows :

Theorem 5.3 : *Let $\mathscr{P}_{\text{II}}$ be a product family of distributions as described in the hypothesis of Theorem* 5.1. *In addition, let the nonregular points of Ω' constitute a set of Lebesgue measure* 0.

Then, conditions (i), (ii) *and* (iii) *of Theorem* 5.1, *for some $r \leqslant s$, are sufficient that there exists a statistic which is sufficient for $\mathscr{P}_{\text{II}}$ in Ω' and is Euclidean of dimension $s < n$ in Ω', and which is continuously differentiable about some regular point of Ω'.*

In fact, in this case, there exists a statistic which is sufficient for $\mathscr{P}_{\text{II}}$ in Ω' and is Euclidean of dimension r in Ω', and which is regular at and analytic about almost every point of Ω'.

Proof : According to the last statement in Theorem 5.1, we have $r = \rho(x)$ for every $x \in \Omega'$, and in particular for every regular point $x^0 \in \Omega'$. According to Theorem 2.4 there is a $\mathscr{P}_{\text{II}}$-sufficient statistic, T^*, which, for almost all (Lebesgue) regular points x^0 of Ω is Euclidean of dimension $\rho(x^0)$ at x^0, and regular at x^0. This statistic is therefore, *a fortiori*, sufficient for $\mathscr{P}_{\text{II}}$ in Ω', is Euclidean of dimension r at almost all regular points of Ω' and is regular at each of these regular points. Moreover, T^* is analytic in a neighbourhood of each of these points, as we see by its construction in Barankin and Katz (1959) from elements of the form of T_0 in Section 4 above, these elements being clearly analytic. Since, under our present hypothesis, the nonregular points of Ω' form a set of Lebesgue measure 0, it follows that T^* is, in fact, Euclidean of dimension r at almost all points of Ω' and regular at and analytic about each of these points. On the Lebesgue null set in Ω' for which this is not true, T^* may be altered, if necessary, to have dimension r, and thus the resulting statistic is, as asserted, sufficient for $\mathscr{P}_{\text{II}}$ in Ω', Euclidean of dimension r in Ω', and regular at and analytic about almost every point of Ω'.

If $r < s$, the statistic just discerned may be augmented by any $s-r$ continuously differentiable components, and thereby we establish, *a fortiori*, the weaker assertion of the present theorem; namely, that there exists a statistic which is sufficient for $\mathscr{P}_{\text{II}}$ in Ω', is Euclidean of dimension $s < n$ in Ω', and is continuously differentiable about some regular point of Ω'.

Theorem 5.3 is therefore proved.

The above three theorems are the global existence theorems that generalize, to the case of nonidentical factor densities, the results of Fisher, Darmois, Koopman and Pitman. We give now the global theorems that generalize, in particular, the Koopman form of results, wherein a specific statistic T is considered. Thus, the next two theorems, as a pair, stand, at the global level, in the same relation to the pair of Theorems 5.1 and 5.2 as Theorem 3.2 stands to Theorem 3.1 at the local level. Of course, Theorem 5.3 yields a sufficient condition for the sufficiency of a given T just as well as Theorem 5.2 does, and the form and proof of this result will be quite evident from our ensuing discussion. (The crucial additional requirement is simply that T^* be almost everywhere a function of T.) However, we shall not here set down explicitly this implication of Theorem 5.3, because it is the sufficient condition drawn rather from Theorem 5.2 that is the natural companion of the necessary condition that derives from Theorem 5.1. This will be clear on the face of it when we have set down our next two theorems; and it will be abundantly verified with the subsequent specialization to the case of identical factor densities.

Theorem 5.4 : *Let $\mathscr{P}_{\Pi}$ be a product family of distributions as described in the hypothesis of Theorem* 5.1. *And let T be a statistic which is Euclidean of dimension $s < n$ in Ω', and is continuously differentiable about some regular point, $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$, of Ω'.*

A necessary condition that T be sufficient for $\mathscr{P}_{\Pi}$ in Ω' is that, for some $r \leqslant s$ and for the regular point x^0, the assertions (i), (ii) *and* (iii) *of Theorem* 5.1 *hold. And furthermore (it is necessary that), if T_0 denotes the statistic defined in the statement of Theorem* 5.2, *relative to the regular point x^0, then T_0 is almost everywhere (Lebesgue) a function of T in Ω'.*

Proof : The first assertion of necessity is an immediate consequence of Theorem 5.1. (The proof of that theorem shows that the regular point x^0 may be pre-assigned, as it is here.) The assertion that T_0 is almost everywhere a function of T in Ω' follows directly from Theorem 2.5 in case $r > 0$. If $r = 0$ the assertion is obvious, since T_0 is then constant over Ω'.

The proof is complete.

Theorem 5.5: *Let $\mathscr{P}_{\Pi}$ be a product family of distributions as described in the hypothesis of Theorem* 5.1. *And let T be a statistic which is Euclidean of dimension $s < n$ in Ω', and is continuously differentiable about some regular point, $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$, of Ω'.*

Sufficient conditions that T be sufficient for $\mathscr{P}_{\Pi}$ in Ω' are as follows :

(1) *for some $r \leqslant s$ and for the given regular point x^0, the conditions* (i), (ii) *and* (iii) *of Theorem* 5.1 *hold;*

(2) *either $r = 0$ or, if $r > 0$ and T_0, N^0 and f^0 are the elements presented in Theorem* 5.2 *(see* (5.3) *and* (5.4)) *for the regular point x^0, then f^0, defined on $\mathscr{R}_{T_0|N^0} \times \Theta$, has a nonvanishing extension f over $\mathscr{R}_{T_0|\Omega} \times \Theta$ such that $f(T_0(x), \theta)$ is analytic over $\Omega' \times \Theta$; and finally*

(3) *(if $r > 0$) T_0 is almost everywhere a function of T in Ω'.*

Proof: According to the first paragraph in the statement of Theorem 5.2, the condition (1) implies immediately that T is sufficient if $r = 0$. If $r > 0$, then, again by Theorem 5.2, conditions (1) and (2) imply that T_0 is sufficient for $\mathcal{P}_{\text{II}}$ in Ω'. Thereupon, condition (3) gives by the usual argument on the factorized form of the family density function, that T is sufficient.

This completes the proof.

We now finally give the specializations of our results to the case of identical factor densities; thus, the next two results are the global counterparts of Corollaries 3.1 and 3.2, respectively. What is to be noted in particular is that we are able to state necessary *and* sufficient conditions in the global case. And the reason for this is that the analytic extension condition set forth in (iv) of Theorem 4.1 is automatically fulfilled in the case of identical factor densities.

Corollary 5.1 : *Let $\mathcal{P}_{\text{II}}$ be the family of distributions defined in* (1.2), *wherein, in particular, the p_i, $i = 1, 2, \ldots, n$, are identically the same function, say p_0. Let Ω_0 be a standard copy of the identical open sets $\Omega_1, \Omega_2, \ldots, \Omega_n$, so that $p_0(\xi, \theta)$ is defined on $\Omega_0 \times \Theta$. Let Ω_0' be a subset of Ω_0 (possibly Ω_0 itself), which is an open, connected set (i.e., an open interval) and is such that p_0 is analytic in $\Omega_0' \times \Theta$. Let Ω' be the n-fold direct product of Ω_0 with itself ; thus, $p(x, \theta) \equiv p_0(x_1, \theta) \cdot p_0(x_2, \theta) \cdot \ldots \cdot p_0(x_n, \theta)$ is analytic in $\Omega' \times \Theta$.*

A necessary and sufficient condition that there exists a statistic which is sufficient for $\mathcal{P}_{\text{II}}$ in Ω' and is Euclidean of dimension $s < n$ in Ω', and which is continuously differentiable about some regular point of Ω', is that for some integer $r \leqslant s$ the following be true :

(i) *p_0 is of exponential type in $\Omega_0' \times \Theta$, as follows ;*

$$\log p_0(\xi, \theta) = b_0(\theta) + \psi_0(\xi) + \sum_{\lambda=1}^{r} b_\lambda(\theta)\psi_\lambda(\xi), \ (\xi, \theta) \in \Omega_0' \times \Theta, \qquad \ldots \quad (5.6)$$

with all functions appearing in this expression being analytic in $\Omega_0' \times \Theta$;

(ii) *there is no constant linear relation among the r functions b_λ, $\lambda = 1, 2, \ldots, r$, in* (5.6); *that is, the expression on the right in* (5.6) *is in reduced form;*

(iii) *for a suitable representation* (5.6) *of p_0, the functions b_λ, $\lambda = 1, 2, \ldots, r$, are constant linear combinations of the logarithmic derivative of p_0 evaluated at (say) the first r coordinates, $x_1^0, x_2^0, \ldots, x_r^0$, of some regular point x^0 in Ω', thus :*

$$b_\lambda(\theta) = \beta_{\lambda 0} + \sum_{i=1}^{r} \beta_{\lambda i} \left(\frac{\partial \log p_0(\xi, \theta)}{\partial \xi} \right)_{\xi = x_i^0}, \quad \lambda = 1, 2, \ldots, r, \qquad \ldots \quad (5.7)$$

wherein the matrix $\|\beta_{\lambda i}\|$ is nonsingular.

For $r = 0$ these three conditions are to be understood as follows : (5.7) *is vacuous and the r-term sum on the right-hand side of* (5.6) *is* 0.

The integer r for which the above properties are verified is unique and is precisely the constant value of the function ρ in Ω'.

Proof: The necessity of these conditions, together with the last statement of the theorem, follows immediately by specializing Theorem 5.1 to identical factor densities.

To prove the sufficiency of the conditions, notice to begin with that, since the present conditions are those of Theorem 5.1 for identical factor densities, the hypothesis in the first paragraph of Theorem 5.2 is satisfied. Therefore, if $r = 0$, the sufficiency of the above three conditions is trivial. Consider, then, $r > 0$. Ready calculations on (5.6) give, in the present instance, the following more explicit form of (5.3), in terms of the suitable j_k's and $\theta^{(k)}$'s affirmed by Theorem 5.2 :

$$\eta_k(x) = n\left(\frac{\partial b_0}{\partial \theta j_k}\right)_{\theta^{(k)}} + \sum_{\lambda=1}^{r}\left(\frac{\partial b_\lambda}{\partial \theta j_k}\right)_{\theta^{(k)}}\left(\sum_{i=1}^{n}\psi_\lambda(x_i)\right) k = 1, 2, \ldots, r \qquad \ldots \quad (5.8)$$

and we have also from (5.6) :

$$\log p(x, \theta) = nb_0(\theta) + \sum_{i=1}^{n} \psi_0(x_i) + \sum_{\lambda=1}^{r} b_\lambda(\theta) \cdot \left(\sum_{i=1}^{n} \psi_\lambda(x_i)\right). \qquad \ldots \quad (5.9)$$

Now, the suitable j_k's and $\theta^{(k)}$'s being what they are (see Section 4), it follows that the matrix

$$\left\| \left(\frac{\partial b_\lambda}{\partial \theta j_k}\right)_{\theta^{(k)}} \cdot \right\|_{\lambda, k = 1, 2, \ldots, r} \qquad \ldots \quad (5.10)$$

is nonsingular. Therefore, the equations (5.8) may be solved for the functions $\sum_{i=1}^{n} \psi_\lambda(x_i)$, $\lambda = 1, 2, \ldots, r$ in terms of the functions η_k, $k = 1, 2, \ldots, r$ (and this solution is valid over all of Ω'). On substituting these solutions into (5.9) we see that, T_0 being as defined in Theorem 5.2, there is a function F such that

$$nb_0(\theta) + \sum_{\lambda=1}^{r} b_\lambda(\theta) \cdot \left(\sum_{i=1}^{n} \psi_\lambda(x_i)\right) = F(T_0(x), \theta), \quad (x, \theta) \in \Omega' \times \Theta. \qquad \ldots \quad (5.11)$$

And if we set $$e^{F(T_0(x), \theta)} = f(T_0(x), \theta) \; ; \; e^{\sum_{i=1}^{n} \psi_0(x_i)} = g(x), \qquad \ldots \quad (5.12)$$

then we have $$p(x, \theta) = f(T_0(x), \theta) g(x), \qquad (x, \theta) \in \Omega' \times \Theta. \qquad \ldots \quad (5.13)$$

From our hypothesis we now see by (5.11) that $F(T_0(x), \theta)$ is analytic over $\Omega' \times \Theta$, and therefore the same is true of $f(T_0(x), \theta)$. This function f, defined over $\mathcal{R}_{T_0|\Omega'} \times \Theta$, is a nonvanishing extension of its restriction to $\mathcal{R}_{T_0|N^0} \times \Theta$, this latter providing the f^0 given in Theorem 5.2. Hence, we have verified that the analytic extension condition in the last paragraph of Theorem 5.2 holds in the present identical-factor situation. Consequently we have forthwith, by the last statement of Theorem 5.2, that there exists a statistic which is sufficient for $\mathcal{P}_{\Pi I}$ in Ω', is Euclidean of dimension $s < n$ in Ω' and is continuously differentiable about some regular point of Ω'.

This concludes the proof that our conditions are, as asserted, sufficient; and therewith the corollary.

Corollary 5.2 : *Let $\mathcal{P}_{\Pi}$ be an identical-factor, product family of distributions as described in the hypothesis of Corollary* 5.1. *And let T be a statistic which is Euclidean of dimension $s < n$ in Ω' and continuously differentiable about some regular point, x^0, of Ω'. Let T_0 again denote the statistic defined in Theorem* 5.2.

A necessary and sufficient condition that T be a sufficient statistic for $\mathscr{P}_{\Pi I}$ in Ω' is that the following be true :

(1) *the conditions* (i), (ii) *and* (iii) *of Corollary* 5.1 *hold, and*

(2) T_0 *is almost everywhere a function of T in Ω'. Or, equivalently, each of the functions* (*see* (5.6))

$$\sum_{i=1}^{n} \psi_\lambda(x_i), \qquad \lambda = 1, 2, \ldots, r \qquad \ldots \quad (5.14)$$

is almost everywhere a function of T in Ω'.

Proof: The equivalence noted in the last statement here is a consequence of the discussion in the proof of Corollary 5.1, which showed that T_0 and the vector function with components (5.14) are functions of each other in Ω'.

If (1) above holds, then the proof of Corollary 5.1 shows that conditions (1) and (2) of Theorem 5.5 are verified. And condition (2) above is identical with condition (3) of that theorem. Hence, Theorem 5.5 establishes the sufficiency of the present conditions (1) and (2).

Conversely, if T is sufficient in Ω', then the present condition (1) holds by virtue of Corollary 5.1. And then condition (2) above is established by Theorem 5.4.

This completes the proof of Corollary 5.2.

This last corollary is Koopman's type of result. However, it differs in its specifics from the actual statements of Koopman. As we have already remarked earlier, the definition of a sufficient statistic that we are working with is the modern, less restrictive one. In these circumstances, allowing ourselves the continuous differentiability of T at some regular point of Ω', instead of adhering to the class of continuous statistics, we are able to obtain necessary *and* sufficient conditions for the sufficiency of T. Koopman was obliged to state two separate theorems, one giving necessary conditions and the other giving sufficient conditions.

References

BAHADUR, R. R. (1954) : Sufficiency and statistical decision functions. *Ann. Math. Stat.*, **25**, 423–462.

BARANKIN, E. W. and KATZ, M. L., Jr. (1959) : Sufficient statistics of minimal dimension. *Sankhyā*, **21**, 217–246.

BARANKIN, E. W. (1960a) : Application to exponential families of the solution of the minimal dimensionality problem for sufficient statistics. *Bull. de l'Inst. Intern. de Stat.*, **38**, (Actes de la 32 Session, Tokyo, 1960), 141–150.

——— (1960b) : Sufficient parameters: solution of the minimal dimensionality problem. *Ann. Inst. Stat. Math.*, Tokyo, **XII**, 91-118.

——— (1961) : A note on functional minimality of sufficient statistics. *Sankhyā*, Series A, **23**, 401–404.

DARMOIS, G. (1935) : Sur les lois de probabilité à estimation exhaustive. *C. R. de l'Acad. des Sc. de Paris*, **2003**, 1265.

DIEUDONNÉ, J. (1960) : *Foundations of Modern Analysis*, Academic Press, New York.

Fisher, R. A. (1922) : On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc. London*, **222** (A), 309–368.

——— (1934) : Two new properties of mathematical likelihood. *Proc. Roy. Soc. London*, **144**(A), 285–307.

KOOPMAN, B. O. (1936) : On distributions admitting a sufficient statistic. *Trans. Amer. Math. Soc.*, **39**, 399–409.

PITMAN, E. J. G. (1936) : Sufficient statistics and intrinsic accuracy. *Proc. Camb. Phil. Soc.*, **32**, 567–579.

Paper received : July, 1962.

STATISTICAL ESTIMATION OF DENSITY FUNCTIONS*

By M. S. BARTLETT

University of Manchester

SUMMARY. The optimum choice of weighting function for smoothing sample density functions is discussed in the cases : (i) probability densities ; (ii) spectral densities. It is shown that the bias contribution to the mean-square error can in principle be eliminated to any required order, though the resulting theoretical gain in efficiency may not be realised except for very large samples. The relation with recent work by Rosenblatt, Daniels and Parzen is noted.

The further problem of estimating the spectra of stationary point processes is also considered.

1. Various important examples of the estimation of density functions arise in statistics, for example, (i) when a sample of independent observations is available from a distribution with a continuous density $f(x)$, where

$$\int_{-\infty}^{\infty} f(x)dx = 1;$$

(ii) when a time-series realisation of extent T is available from a (real) stationary stochastic process $X(t)$ with continuous spectral density $g(\omega) = \sigma_x^2 f_+(\omega)$, where

$$\int_0^{\infty} f_+(\omega)\, d\omega = 1.$$

For simplicity, we shall in place of (ii) consider the analogous problem for a discrete-time realisation of extent $T = n$ with unit intervals, so that $f_+(\omega)$ is defined up to $\omega = \pi$, and analogously for (i), assume that $f(x)$ is zero outside a finite interval of known extent.

The second problem has received most attention in the literature, but the comparability of the first problem was emphasized in an interesting paper by Rosenblatt (1956). It is convenient to discuss problem (i) first, as in some ways simpler than (ii), and with this object in view we recapitulate below some of the relevant formulae. Finally, after returning to (ii), we extend the technique to cover the spectra of auto-covariance densities.

Let our estimate of a probability density $f(x)$ at x be denoted by the functional

$$g[x;\ w_x(u)] = \int w_x(u)\ dF_s(u).$$

where $F_s(u)$ is the *sample* cumulative distribution function obtained from n independent observations.

*Presented in summarized form at the International Congress of Mathematicians, Stockholm, 1962.

Then we readily obtain

$$E[g] = \int w_x(u) f(u) du,$$
$$\text{var}\,[g] = \int\int w_x(u)\ w_x(v)\ \text{cov}\,[dF_s(u), dF_s(v)]$$
$$= \frac{1}{n}[\int w_x^2(u) f(u) du - [\int w_x(u) f(u)\, du]^2].$$

The mean square error S is given by

$$\text{var}\,[g] + [E\{g\} - f(x)]^2 = \frac{1}{n}[\int w_x^2(u) f(u) du$$
$$-[\int w_x(u) f(u) du]\,]^2 + [\int w_x(u) f(u) du - f(x)]^2. \quad \ldots \quad (1)$$

In order to study the behaviour of (i) further, Rosenblatt considers the case where $f(u)$ is well-behaved near x, and can be written

$$f(u) = f(x) + (u-x) f'(x) + \frac{1}{2}(u-x)^2 f''(x) + 0[(u-x)^3].$$

Suppose also that for x inside the range of u, we have

$$w_x(u) = w(x-u)$$

where $w(u)$ is an even function of u. Then if $w(u)$ is zero (or effectively zero) for $|u| > h$, where h is small, and we suppose further that

$$\int w(u) du = 1,$$

we have the first term on the right-hand side of (1) of order $1/(hn)$. Then if

$$\int u^2 w(u) du = v.$$
$$\int w^2(u)\, du = W,$$

the dominant terms in (1) (for $v \neq 0$) are

$$S \sim \frac{1}{n} W f(x) + \frac{1}{4} v^2 [f''(x)]^2. \quad \ldots \quad (2)$$

For example, if $w(u)$ is $1/(2h)$ between $-h$ and h, we have $W = 1/(2h)$, $v = \frac{1}{3}h^2$, and (2) becomes

$$S \sim \frac{f(x)}{2hn} + \frac{h^4}{36}[f''(x)]^2. \quad \ldots \quad (3)$$

Rosenblatt notes that if we put $h = cn^{-\alpha}$, α should for large n be taken to be $\frac{1}{5}$,

and
$$C^5 = \frac{9f(x)}{2[f''(x)]^2},$$

when (3) becomes
$$S \sim \frac{5}{36}\left(\frac{9}{2}\right)^{4/5} f^{4/5} (f'')^{2/5} n^{-4/5}. \quad \ldots \quad (4)$$

2. However, if under the conditions for which (2) is valid we consider $w(u)$ to be an arbitrary (and not necessarily, as Rosenblatt assumed, a non-negative) function which is to be optimised, it is readily found (for example, by the method used for an analogous problem in Bartlett and Medhi (1955)) that it should be of the form

$$w(u) = C(1-u^2/a^2).$$

Now by choice of a, we can here make v zero; we find $u = h\sqrt{\frac{3}{5}}$, whence $C = W = 9/(8h)$, and

$$S \sim \frac{9f}{8hn}, \qquad \dots \quad (5)$$

a formula in which, by taking h small but fixed, we can apparently restore the result $S = 0(n^{-1})$. Of course, this is not precisely true, for we must now include the next order term in the bias

$$\frac{1}{24} f^{(iv)}(x) \int u^4\, w(u)du.$$

The 'fourth moment' of $w(u)$, V, say, is in the above case $9h^4/110$, whence more accurately

$$S \sim \frac{9f}{8hn} + \left(\frac{1}{24} f^{(iv)} \cdot \frac{9h^4}{110}\right)^2. \qquad \dots \quad (6)$$

After inclusion of this term, however, we should go back and modify our optimum function $w(u)$ to arrange for V to be zero as well as v, and so on indefinitely. In this way, the result $S = 0(n^{-1})$ can be approached.*

It seems doubtful whether such a repeated correction would be practically superior over the first correction given by

$$w(u) = \frac{9}{8h}\left(1 - \frac{5u^2}{3h^2}\right), \qquad \dots \quad (7)$$

which we will compare with the rectangular weighting function of Section 1. Let us consider, for example, the estimation of the mode of the normal distribution with density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}x^2/\sigma^2}$$

for which we note that

$$f(0) = \frac{1}{\sqrt{2\pi}\sigma}, \quad f''(0) = -\frac{1}{\sqrt{2\pi}\sigma^3}, \quad f^{(iv)}(0) = \frac{3}{\sqrt{2\pi}\sigma^5}$$

*The technique suggested here appears related to an independent and more mathematical discussion by Parzen (1961), so I might perhaps note that it was first mentioned in a lecture course on time series given at University College, London, in the academic year 1960-61. The effective use of orthogonal polynomial weighting functions should also be compared with the procedure for estimating spectral densities proposed by Daniels (1962) who, however, relates orthogonal polynomials with the density function rather than with the weighting function. For rather a different approach using prior distributions, see Whittle (1958). Cf. also the paper by Priestley (1962).

Thus we choose $h = (4.5/n)^{1/5}\sigma$, and

$$\sqrt{S}/f(0) \sim \frac{5}{8} \Big/ \sqrt{[h\, n\, f(0)]}. \qquad \ldots \quad (8)$$

For this fraction to be 0.1, say, we require $n = 303$. From (6) we then have for comparison

$$\sqrt{S}/f(0) \sim \frac{9}{8} \Big/ \sqrt{[h\, n\, f(0)]}. \qquad \ldots \quad (9)$$

Within limits, h may be taken as large as we please, but for definiteness we choose the same numerical value as in (8) for both h and n, whence the value in (9) becomes 0.134, thus stressing the obvious point that asymptotic behaviour has no relevance to samples of fixed size. However, for fixed h the value in (9) will eventually become smaller than the value in (8) provided h in (8) is taken to be $(4\frac{1}{2}/n)^{1/5}\sigma$. For example, for n 32 times larger, the value in (8) becomes 0.025, whereas that in (9) becomes 0.0237. Of course, in this last case we must check that the neglected contribution in (9) due to $f^{(iv)}(0)$ is in fact still negligible, but this may be easily checked from the more accurate expression (6).

3. It is evident that a similar approach will be possible for the estimation of a spectral density $f_+(\omega)$, the only differences being (a) a slight difference in the functional dependence on $f_+(\omega)$ and (b) the fact that the moment formulae are only asymptotically valid.

We consider for simplicity the related density function

$$\rho(\lambda) = 2\pi\sigma_x^2 f_+(\lambda),$$

and write as our estimate of $\rho(\lambda)$ at λ the quantity

$$g(\lambda; w_\lambda(\omega)) = \int_0^{\Pi} I_x(\omega) w_\lambda(\omega) d\omega, \qquad \ldots \quad (10)$$

where $I_x(\omega)$ is the usual periodogram intensity at the point ω obtained from a sequence of values $X_1, X_2, \ldots, X_n$. It is possible to express (10) to a sufficient approximation equivalently as

$$\frac{2\pi}{n} \Sigma_p \, I_x(\omega_p) \, w_\lambda(\omega_p),$$

where Σ_p denotes summation over $I_x(\omega) w_\lambda(\omega)$ evaluated at the discrete steps $\omega_p = 2\pi p/n$, ($p = 0, 1, \ldots \frac{1}{2}n$, for n even with appropriate connections at 0 and $\frac{1}{2}n$). Then under suitable regularity conditions on $f_+(\omega)$ and $w_\lambda(\omega)$ it is well known that asymptotically

$$E[g(\lambda)] \sim \int_0^{\Pi} \rho(\omega) w_\lambda(\omega) d\omega, \qquad \ldots \quad (11)$$

$$\text{var}\,[g(\lambda)] \sim \frac{2\pi}{n} \int_0^{\Pi} \rho^2(\omega)\, w_\lambda^2(\omega) d\omega, \qquad \ldots \quad (12)$$

whence the mean-square error S is given by

$$\frac{2\pi}{n}\int_0^{\Pi}\rho^2(\omega)w_\lambda^2(\omega)d\omega+\left[\int_0^{\Pi}\rho(\omega)w_\lambda(\omega)\,d\omega-\rho(\lambda)\right]^2. \qquad \ldots \quad (13)$$

For the uniform weighting function $w_\lambda(\omega) = 1/(2h)$, ($\omega$ in $\lambda-h$ to $\lambda+h$), we have similarly to (3) and (4) (cf. Grenander and Rosenblatt, 1957, p. 154)

$$S \sim \frac{\pi\rho^2(\lambda)}{hn}+\frac{h^4}{36}[\rho''(\lambda)]^2 = n^{-4/5}\left[\frac{\pi\rho^2(\lambda)}{c}+\frac{c^4}{36}[\rho''(\lambda)]^2\right] \qquad \ldots \quad (14)$$

for $h = cn^{-4/5}$;

$$= \frac{\frac{5}{4}\,\pi^{4/5}\,[\rho(\lambda)]^{8/5}\,[\rho''(\lambda)]^{2/5}}{9^{1/5}\,n^{4/5}} \qquad \ldots \quad (15)$$

for $c = [9\pi^2[\rho(\lambda)/\rho''(\lambda)]^2]^{1/5}$.

Again, however, by choice of $w_\lambda(\omega)$ we can make the bias vanish to any desired order. Keeping for simplicity to the same function as in (7), we write

$$w_\lambda(\omega) = \frac{9}{8h}\left[1-\frac{5(\omega-\lambda)^2}{3h^2}\right]. \qquad \ldots \quad (16)$$

For a parabolic density law, at least in the neighbourhood of λ, this weighting function will, as for the probability density estimation problem, give a mean-square error decreasing ultimately as $1/n$, in contrast with the $n^{-4/5}$ of formula (15). For example, for the parabolic density function

$$\rho(\omega) = \frac{3\pi}{4}\left[1-\frac{(\omega-\lambda)^2}{4}\right],$$

$$\rho''(\lambda) = 3\pi/8, \quad \text{and from (15)}$$

$$\sqrt{S}/\rho(\lambda) \sim 1.013\, n^{-2/5}. \qquad \ldots \quad (17)$$

For this fraction to be 0.1, we require $n = 537$. For (16), we have then (for the same value of h) a fraction 0.134, a value greater than 0.1. Nevertheless, for $n = 32\times 537$, the value in (17) is 0.025, while the weighting function (16) for fixed h gives now a value for $\sqrt{S}/\rho(\lambda)$ of 0.0237, analogously to the probability density case.

4. If $I_x(\omega)$ is available at integral values of p, where $\omega_p = 2\pi p/n$, then the weighting function (16) becomes only approximately correct, and the exact values appropriate to the number of discrete points in the interval h to h may be substituted. These are readily found from the analogous conditions that the sum of the weights is unity, and the orthogonality with $(\omega-\lambda)^2$ preserved. For example, for 8 points in the interval the weights are:—

$$-\frac{3}{32} \quad \frac{3}{32} \quad \frac{7}{32} \quad \frac{9}{32} \quad \frac{9}{32} \quad \frac{7}{32} \quad \frac{3}{32} \quad -\frac{3}{32}$$

and for 16 points in the interval:

$$\frac{-91}{1344} \quad \frac{-21}{1344} \quad \frac{39}{1344} \quad \frac{89}{1344} \quad \frac{129}{1344} \quad \frac{159}{1344} \quad \frac{179}{1344} \quad \frac{189}{1344} \ldots$$

It is recalled that in the case of estimating spectral densities various formulae have been proposed in terms of the autocovariance or autocorrelation function, the main purpose (at least by myself) being to save computation. Now, however, that large-scale computers are often available, there is much to be said for working directly with the periodogram. The modified estimate discussed above will then sometimes be of interest, especially if it is important to reduce the effect of bias; but the value of the uniform weighting function, as first proposed in this context by Daniell (1946), should also be emphasized. It is especially efficient for testing departure from a uniform spectrum, giving rise in this case to sampling quantities distributed asymptotically as χ^2's.

5. Further examples of density functions arise in the theory of point processes, for which 'product density' or 'factorial-moment density' functions are defined (see Bartlett, 1955, Section 3.42) and may require estimation. For example, in the case of a stationary point process $dN(t)$, the second-order density function defines a 'covariance density' function

$$\mu(\tau) = \frac{E[dN(t+\tau)dN(t)]}{(dt)^2} - \lambda^2, \ (\tau > 0), \qquad \dots \quad (18)$$

where

$$\lambda = \frac{E[dN(t)]}{dt}, \qquad \dots \quad (19)$$

analogously to the covariance function $W(\tau)$ of a stationary process $X(t)$. The sampling properties of an estimate of $\mu(\tau)$ can, like those of an estimate of $W(\tau)$, be investigated, but are rather complicated; and it seems more convenient to discuss the estimation of the Fourier transform of $\mu(\tau)$, which is the analogue of the spectrum corresponding to $W(\tau)$ (see Bartlett, 1955a, Section 6.12).

It will be shown that periodogram intensities $I_N(\omega)$ may be defined with similar asymptotic properties to those of $I_x(\omega)$, so that the smoothing technique developed in Section 3 above will be readily applicable to this further problem.

We define

$$g(\omega) = \int_{-\infty}^{\infty} e^{-i\tau\omega}\, \mu(\tau) d\tau. \qquad \dots \quad (20)$$

Note that $\mu(\tau)$ is only defined in (18) above for $\tau > 0$. It is a symmetric function, but for $\tau = 0$ we have

$$E[(dN(t))^2] = E[dN(t)] = \lambda\ dt \qquad \dots \quad (21)$$

so that the integral (20) has a contribution λ at $\tau = 0$ and $g(\omega)$ is of the form

$$g(\omega) = \lambda + g_\mu(\omega). \qquad \dots \quad (22)$$

For ω defined for non-negative values only, we write

$$g_+(\omega) = 2\lambda + 2g_\mu(\omega). \qquad \dots \quad (23)$$

For events at random times $T_1, T_2, \ldots, T_n$ in the interval $(0, T)$ we define

$$J(\omega) = \sqrt{\frac{2}{T}} \sum_{s=1}^{n} e^{iT_s\omega} = \sqrt{\frac{2}{T}} \int_0^T e^{it\omega}\, dN(t). \qquad \ldots \quad (24)$$

We have, analogously to the properties of an ordinary periodogram intensity $I_x(\omega)$

$$E[I(\omega)] = E[J(\omega)J^*(\omega)]$$

$$= \frac{2}{T}\left[\int_0^T \lambda dt + \int_0^T \int_0^T e^{i(u-v)\omega}\, \mu(u-v) du\, dv\right] \quad \ldots \quad (25)$$

$$\sim g_+(\omega) \text{ for large } T.$$

To investigate the sampling properties of $I_x(\omega)$ it is usual (e.g. Bartlett, 1955a, Section 9.2) to assume that $X(t)$ is a 'linear process.' The analogous interpretation of the point process $N(t)$ will be that it can be regarded as a Poisson process with rate Λ which is itself a stationary linear process $\Lambda(t)$ with mean λ. It then follows (Bartlett, 1955b) that the relation between the characteristic functional of $N(t)$ and $\Lambda(t)$ is

$$E[\exp\, i \int_0^T \theta(t)\, dN(t)] = E_\Lambda\, [\exp \int_0^T \Lambda(t)[e^{i\theta(t)}-1]dt]. \qquad \ldots \quad (26)$$

This relation includes for example the result contained in the equation leading to (25)

$$E[I(\omega)] = 2\lambda + E[I_\Lambda(\omega)], \qquad \ldots \quad (27)$$

where

$$I_\Lambda(\omega) = J_\Lambda(\omega) J^*_\Lambda(\omega),$$

$$J_\Lambda(\omega) = \sqrt{\frac{2}{T}} \int_0^T \Lambda(t)\, e^{it\omega}\, dt,$$

and $\mu(\tau)$ is now also the autocovariance function of $\Lambda(t)$.

In (26) choose

$$i\theta(t) = \sqrt{(2/t)}(\theta_1\, e^{it\omega} + \theta_2\, e^{-it\omega}).$$

Then by expanding $e^{i\theta(t)}$ on the right-hand side of (26) we find

$$E[\exp[\theta_2 J(\omega) + \theta_2 J^*(\omega)]] = E_\Lambda\, [\exp[\theta_1 J_\Lambda(\omega) + \theta_2 J^*_\Lambda(\omega) + 2\lambda\theta_1\theta_2]] + 0(1/\sqrt{T})$$

or, by taking logarithms to obtain cumulant functions,

$$K(\theta_1, \theta_2) \sim K_\Lambda(\theta_1, \theta_2) + 2\lambda\theta_1\theta_2, \qquad \ldots \quad (28)$$

where the cumulant function on the left refers to $J(\omega)$, $J^*(\omega)$, and that on the right to $J_\Lambda(\omega)$, $J^*{}_\Lambda(\omega)$.

The use of complex quantities may obscure the interpretations of the extra term $2\lambda\theta_1\theta_2$ in (28), but this result implies that $J(\omega)$ is asymptotically equivalent to $J_\Lambda(\omega)$ apart from the addition of an independent (complex) component with real and

imaginary parts each uncorrelated normal with zero mean and variance λ. We deduce the following asymptotic results in addition to (25).

$$\begin{aligned} \text{var } [I(\omega)] &= E[I^2(\omega)]-[E[I(\omega)]]^2 \\ &= E[J^2(\omega)J^{*2}(\omega)]-[E[I(\omega)]]^2 \\ &\sim \kappa_{22}+\kappa_{20}\kappa_{02}+\kappa_{11}^2 \\ &\sim \kappa_{11}^2 \sim [E[I(\omega)]]^2 \qquad \ldots \quad (29) \end{aligned}$$

where these cumulants relate to $K(\theta_1, \theta_2)$ and it is noted that $E(J(\omega)] \sim 0$ for $\omega \neq 0$. The last line then follows because it is known to be true for $J_\wedge(\omega)$ and $I_\wedge(\omega)$; from (28) K_{22}, K_{20} and K_{02} are also zero for the extra component, whose cumulants are additive to those to $J_\wedge(\omega)$.

Notice that the extra component adds to the variance of $I(\omega)$ as well as to its mean, so that the fluctuations of $I(\omega)$ are similar to those in standard periodogram analysis. We may consider

$$I'(\omega) = I(\omega)-2N(T)/T \qquad \ldots \quad (30)$$

where the second term corrects for the extra component by means of a sample estimate, but the variance is *not* thereby reduced.

That we may regard $I(\omega)$ as having similar asymptotic properties to $I_\wedge(\omega)$ has only been shown above at one point ω, and in particular for the variance; but it may similarly be shown for two points, and the covariance of $I(\omega)$ and $I(\omega')(\omega, \omega' \neq 0)$, by writing in (26)

$$i\theta(t) = \sqrt{\frac{2}{T}}(\theta_1 e^{it\omega}+\theta_2 e^{-it\omega}+\theta_1 e^{it\omega'}+\theta_2 e^{-it\omega'}).$$

It should perhaps be added that the asymptotic behaviour of $I(\omega)$ for large T for a particular ω, or pair $\omega \,.\, \omega'$ could be deduced similarly for any finite set of separate ω's, but could not be expected to hold for a set of increasing number without further conditions. This limitation applies in the standard periodogram case, but as ω is unrestricted in range raises some further queries. For example, for a particular set of observations, $t_0, t_1, \ldots, t_N$ some effective upper limit on the range of ω is required, but no precise discussion of this point is attempted (cf. also the statistical analysis in Section 6).

6. To illustrate the technique of Section 5, two examples were analysed. (I) The first set of data consists of times at which 129 successive vehicles passed a point on a road;* (II) the second is an artificial realization of a purely random or Poisson process, with a mean interval of unity.

*I am indebted for these figures to Dr. A. J. Miller, who had supplied them as a class example for demonstrating other types of statistical analysis I have suggested for such data. The programming of the computer calculations is due to Mr. D. Walley.

In Example I the total time elapsing is 20235 (in units of 1/10 second), giving an average interval of 158.1. This was reduced to the order of unity by taking 160 as unit time interval. Thus if the recorded times are $t_0, t_1, \ldots, t_{129}$, the scaled time-intervals

$$T_i = (t_i - t_0)/160, \ (i = 1, \ldots, 128),$$

were used, and $J(\omega_p)$ defined as

$$J(\omega_p) = A(\omega_p) + iB(\omega_p) = \left(\sqrt{\frac{2}{T_{128}}}\right) \sum_{s=1}^{n} e^{iT_s \omega p}.$$

The values of ω_p chosen were of the form $2\pi p/N$, where N and p are integers. In standard periodogram analysis, N would be of order n, where n is the sample length, and was thus chosen to be 128. Moreover, the range of p, when $N \sim n$, is $\frac{1}{2}n\,(\omega_p \leqslant \pi)$. Correspondingly, in the present type of problem, we should take the range of p to be at least $\frac{1}{2}N$, or 64. It was taken to be 4 times this amount, as a reasonable compromise between taking too many values of p and taking a high enough maximum for the bulk of the relevant variation of the spectrum to be included.

In Example II the artificial series also consisted of 128 time-intervals, but these did not require any preliminary scaling. The values of N and the range of p were taken as for Example I. A more detailed record and discussion of this analysis will be given elsewhere, but a summary of the analysis is provided by Tables 1 and 2. These provide smoothed $I(\omega)$ estimates for each example (i) using a uniform weighting function over 16 points (ii) using a parabolic weighting function over 16 points (with weights as given in Section 4). It should of course be remembered that in Example II the spectral function is known to be constant (with mean 2), and the uniform weighting function would necessarily be superior to the parabolic, as is suggested by the results. In Example I it was known by previous analyses that the process was not Poisson. This would be tested on the present approach with the uniform weighting function (for which the smoothed $I(\omega)$ are approximately distributed proportionally to χ^2 with 32 degrees of freedom); but once the non-constancy of the spectral function is established, the use of the parabolic weighting function might be preferred as giving a more unbiased, and under certain conditions more accurate, estimate of the true function. The value of $2N(T)/T$ should be noted in this case viz. 2.024.

TABLE 1. ESTIMATED SPECTRAL VALUES FOR TRAFFIC DATA

	(i)	(ii)		(i)	(ii)		(i)	(ii)		(i)	(ii)
1	4.702	5.970	5	2.974	1.761	9	2.136	2.478	13	2.128	3.086
2	3.120	2.790	6	2.353	3.139	10	2.029	1.246	14	1.118	0.590
3	4.380	2.876	7	2.706	2.467	11	1.701	1.879	15	1.163	1.495
4	2.761	2.628	8	3.343	3.547	12	2.477	1.597	16	1.788	1.507

TABLE 2. ESTIMATED SPECTRAL VALUES FOR ARTIFICIAL SERIES

	(i)	(ii)		(i)	(ii)		(i)	(ii)		(i)	(ii)
1	1.740	1.497	5	1.647	1.823	9	1.598	1.686	13	2.038	1.642
2	2.561	3.425	6	2.146	2.573	10	1.734	2.072	14	2.202	2.909
3	1.606	1.101	7	1.499	1.597	11	1.980	1.958	15	1.779	0.789
4	3.323	4.027	8	1.621	1.212	12	0.708	0.662	16	1.699	1.293

REFERENCES

BARTLETT, M. S. (1955a) : *An Introduction to Stochastic Processes*, (Cambridge).

——— (1955b) : Contribution to the discussion on the paper by D. R. Cox, Some statistical methods connected with series of events. *J. Roy. Stat. Soc.*, B. **17**, 129.

BARTLETT, M. S. and MEDHI, J. (1955) : On the efficiency of procedures for smoothing periodograms for time series with continuous spectra. *Biometrika*, **42**, 143.

DANIELL, P. J. (1946) : Contribution to the discussion at the *Symposium on Autocorrelation in Time-Series*. *J. Roy. Stat. Soc.*, B. **8**, 27.

DANIELS, H. E. (1962) : The estimation of spectral densities. *J. Roy. Stat. Soc.*, B, **24**, 185.

GRENANDER, U. and ROSENBLATT, M. (1957) : *Statistical Analysis of Stationary Time Series*, Wiley, New York.

PARZEN, E. (1961) : On estimation of a probability density function and mode. O.N.R. Technical Report No. 40.

PRIESTLEY, M. B. (1962) : Basic considerations in the estimation of spectra. *Technometrics*, **4**, 551.

ROSENBLATT, M. (1956) : Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.*, **27**, 832.

WHITTLE, P. (1958) : On the smoothing of probability density functions. *J. Roy. Stat. Soc.*, B, **20**, 334.

Paper received : April, 1963.

SOME LIMIT THEOREMS FOR THE DODGE-ROMIG AOQL SINGLE SAMPLING INSPECTION PLANS*

By A. HALD and E. KOUSGAARD

University of Copenhagen

1. Introduction

In a previous paper by Hald (1962) limit theorems for the Dodge-Romig LTPD single sampling inspection plans have been derived. The purpose of the present paper is to find similar results for the AOQL plans. Dodge and Romig (1941) did not consider this problem in their paper where they derived the equations for the AOQL plans and tabulated such plans for lot sizes up to 100.000. Both from theoretical and practical points of view it is, however, interesting to develop an explicit asymptotic solution to Dodge and Romig's equations and study to what extent this solution is valid for finite lots.

The main results are that the highest allowable fraction defective in the sample converges to the AOQL, the difference being of order $\sqrt{\log n}/\sqrt{n}$ (Theorem 2), and that sample size asymptotically is proportional to the logarithm of lot size (Theorem 3). It is further shown that the producers risk asymptotically decreases inversely proportional to lot size and that the average amount of inspection for lots of process average quality apart from sampling inspection is independent of lot size. Finally, numerical investigations have shown that the asymptotic formulas for acceptance number and sample size are good approximations to the exact solution also for small lot sizes and a compact graphical representation of the asymptotic solution is given.

From a purely probabilistic point of view the most interesting is perhaps the result regarding the Poisson distribution stated in Theorem 1.

2. Relation between the exact and asymptotic solutions

The notation has been kept as close as practical to Dodge and Romig's. The given parameters are the average outgoing quality limit, p_L, the process average fraction defective, $\bar{p}$, and the number of items in the lot, N. The problem is to determine the number of items in the sample, n, and the acceptance number, c, from the following two requirements : (1) The maximum value of the average fraction defective after sampling inspection with total inspection of rejected lots, replacing all defective items found by good ones, shall be equal to p_L. (2) The average number of items inspected per lot of process average quality shall be a minimum, assuming that the remainder of rejected lots is inspected.

---

* Prepared in part at the Statistical Techniques Research Group, Princeton University, with support from the Office of Naval Research, and under ONR Contract Number N62558-3073.

For incoming lots of size N from a binomially controlled process with fraction defective equal to p the average probability of acceptance is

$$B(c, n, p) = \sum_{x=0}^{c} \binom{n}{x} p^x q^{n-x}$$

and the average fraction defective after inspection becomes approximately

$$p_A = \left(1-\frac{n}{N}\right) pB(c, n, p). \quad \ldots \quad (1a)$$

The approximation involved consists in replacing the average number of defectives found in accepted lots by np which is usually a negligible error.

For $p_L \leqslant 0.10$ Dodge and Romig have used the cumulative Poisson distribution, $B(c, np)$, instead of the binomial, i.e. (1a) is replaced by

$$p_A = \left(1-\frac{n}{N}\right) pB(c, np). \quad \ldots \quad (1b)$$

The value of p maximizing p_A is found from the equation $dp_A/dp = 0$ which leads to

$$B(c, x) = xb(c, x), \qquad x = np_1, \quad \ldots \quad (2)$$

where $b(c, x) = e^{-x} x^c/c!$ denotes the Poisson probability and $\max\limits_{p} p_A = p_L$ for $p = p_1$.

Introducing the auxiliary variables $m = np_L$ and $M = Np_L$ we find from (1b) for $p = p_1$

$$m = \left(1-\frac{m}{M}\right) xB(c, x). \quad \ldots \quad (3)$$

Equation (2) determines a relation between x and c which may be used to eliminate x from (3). Thus, from (3) we may find m as a function of c and M, $m = m_{c,M}$ say.

The average number of items inspected per lot of process average quality is

$$I(\bar{p}) = n+(N-n)(1-B(c, n, \bar{p})). \quad \ldots \quad (4a)$$

Approximating the binomial by the corresponding Poisson distribution and introducing $z = I(\bar{p})p_L$ and $r = \bar{p}/p_L$ we obtain

$$z = m+(M-m)(1-B(c, rm)) = M-(M-m)B(c, rm) \quad \ldots \quad (4b)$$

which is proportional to the function minimized by Dodge and Romig for $r < 1$ and $p_L \leqslant 0.10$.

It should be noted that m, M, z, and r here are defined similarly as in the paper by Hald (1962) with the natural modification that p_t has been replaced by p_L.

The problem of minimizing z with respect to c when $m = m_{c,M}$ is formally the same as the one treated by Hald (1962). The upper boundary of the zone in the (M, r) plane with acceptance number c is therefore given by the equation

$$M = \frac{m_{c+1,M}\, B(c+1,\, rm_{c+1,M}) - m_{c,M}\, B(c,\, rm_{c,M})}{B(c+1,\, rm_{c+1,M}) - B(c,\, rm_{c,M})} = M(c), \qquad \dots \quad (5)$$

which means that for all M in the interval $(M(c-1), M(c))$ we have $z(c) < z(c+i)$, $i = \pm 1, \pm 2, \dots$ whereas for $M = M(c)$ we have $z(c) = z(c+1) < z(c+i)$, $i = -1, \pm 2, \pm 3, \dots$.

The asymptotic solution is obtained by treating (c, n, N) as continuous and approximating $B(c, rm)$ by a differentiable function. Replacing the difference equation $\Delta z = 0$ for the determination of $M(c)$ by the differential equation $dz/dc = 0$ we find $M = \bar{M}(c)$ and the corresponding approximate relation $M(c) \simeq \bar{M}(c+\frac{1}{2})$.

Instead of using (2) and (3) for eliminating m from (4b) we might just as well have eliminated c and minimized z with respect to m.

The procedure in the following will be to rewrite (2) as an integral and then find a manageable relation between c and x from an asymptotic expansion of the integral. By means of this relation x is eliminated from (3) and c is found as a function of m which is used to eliminate c from (4b). From $dz/dm = 0$ we finally find an expansion for M in terms of m.

3. Relation between sample size and acceptance number

Theorem 1 : *The equation*

$$B(c, x) = xb(c, x) \qquad \dots \quad (6)$$

has the asymptotic solution

$$c = x + \sqrt{x \log \frac{x}{2\pi}} + \frac{1}{6} \log \frac{x}{2\pi} - \frac{1}{2} + o(1). \qquad \dots \quad (7)$$

Proof: By means of the relation between the cumulative Poisson distribution and the incomplete Gamma function we may rewrite (6) as

$$\frac{1}{c!} \int_x^\infty z^c\, e^{-z}\, dz = \frac{1}{c!}\, e^{-x}\, x^{c+1}.$$

Putting $z = x(1+t)$ we find

$$F(c, x) = \int_0^\infty (1+t)^c e^{-xt}\, dt = 1. \qquad \dots \quad (8)$$

Introducing $c = x(1+\delta)$ we want to determine $\delta = \delta_x$ for $x \to \infty$.

First we show that $\delta \to 0$ but at a slower rate than $1/\sqrt{x}$. Let us suppose that asymptotically $c - x \sim \alpha\sqrt{x}$, i.e. $\delta \sim \alpha/\sqrt{x}$, where α is positive and finite. From the central limit theorem it then follows that the left hand side of (6) tends to $\Phi(\alpha)$ whereas the right hand side is of the order $\phi(\alpha)x$ and thus tends linearly to infinity.

Consequently the assumption $\delta \sim \alpha/\sqrt{x}$ leads to a contradiction. Suppose next that $c-x \sim \alpha x$, i.e. $\delta \sim \alpha$ where α is positive and finite. This means that the standardised normal deviate becomes $(c-x)/\sqrt{x} \sim \alpha\sqrt{x}$. It then follows from a formula by Blackwell and Hodges (1959) that the left hand side of (6) tends exponentially to 1 whereas the right hand side tends exponentially to zero. Thus also this assumption leads to a contradiction. Since the two assumptions, $\delta \sim \alpha/\sqrt{x}$ and $\delta \sim \alpha$, lead to discrepancies between the left and the right hand sides of (6) going in opposite directions we have that $\delta \to 0$ at a slower rate than $1/\sqrt{x}$.

Rewriting (8) we find

$$F(c, x) = \int_0^\infty e^{xg(t)}\, dt \qquad \dots \quad (9)$$

with

$$g(t) = (1+\delta)\log(1+t)-t, \quad t \geqslant 0, \qquad \dots \quad (10)$$

$$g'(t) = \frac{1+\delta}{1+t}-1$$

and

$$g^{(r)}(t) = (-1)^{r-1}(r-1)!(1+\delta)(1+t)^{-r} \text{ for } \quad r \geqslant 2.$$

It follows that $g(t)$ is increasing for $0 \leqslant t \leqslant \delta$ and decreasing for $t > \delta$, that $g(0) = 0$,

$$\max_t g(t) = g(\delta) = (1+\delta)\log(1+\delta)-\delta > 0,$$

$g(t) = 0$ for $t \sim 2\delta$, and $g(t) \to -\infty$ for $t \to \infty$.

Since $\delta \to 0$ the essential contribution to $F(c, x)$ for $x \to \infty$ comes from the neighbourhood of $t = \delta$ where we have

$$xg(t) = xg(\delta)+x\sum_{\nu=2}^{\infty}(-1)^{\nu-1}\frac{(t-\delta)^\nu}{\nu(1+\delta)^{\nu-1}}.$$

Introducing

$$u = (t-\delta)\sqrt{x}/\sqrt{1+\delta} \qquad \dots \quad (11)$$

we find

$$xg(t) = xg(\delta)-\frac{u^2}{2}-\sum_{\nu=3}^{\infty}\frac{(-u)^\nu}{\nu}((1+\delta)x)^{-(\nu/2)+1}. \qquad \dots \quad (12)$$

From the properties of $g(t)$ described above and from the fact that $\delta \to 0$ it follows that it is always possible for any fixed $t = t_0 > 0$ to find $x = x_0$, say, so that $g(t_0) < 0$ for all $x > x_0$ which implies that

$$F(c, x) = \int_0^{t_0} e^{xg(t)}dt(1+O(e^{-x})) \quad \text{for} \quad x > x_0.$$

Changing the variable in the integral from t to u defined by (11) and using (12) we get

$$F(c, x) = \sqrt{\frac{2\pi}{x}(1+\delta)}\, e^{xg(\delta)} \int_{-\delta\sqrt{x}/\sqrt{1+\delta}}^{(t_0-\delta)\sqrt{x}/\sqrt{1+\delta}} \phi(u)du\left(1+O\left(\frac{1}{\sqrt{x}}\right)\right) \qquad \dots \quad (13)$$

where $\phi(u)$ denotes the standardised normal frequency function.

To get a first approximation to δ we observe that for $\delta \to 0$ and $\delta\sqrt{x} \to \infty$ the main terms of $\log F$ becomes

$$\log F(c, x) = -\tfrac{1}{2} \log \frac{x}{2\pi} + \tfrac{1}{2} x\delta^2 + \ldots \qquad \ldots \quad (14)$$

since
$$g(\delta) = \frac{\delta^2}{2} - \frac{\delta^3}{6} + \frac{\delta^4}{12} - \ldots .$$

From (8) we have $\log F = 0$ which leads to

$$\delta \sim \sqrt{\frac{1}{x} \log \frac{x}{2\pi}} . \qquad \ldots \quad (15)$$

By means of this result and Mill's ratio we can evaluate the integral in (13) which becomes asymptotically $1 - \dfrac{1}{x\delta}$.

Expanding the three terms of $\log F$ we finally find

$$\log F(c, x) = -\frac{1}{2} \log \frac{x}{2\pi} + \frac{1}{2}\delta + \frac{1}{2} x\delta^2 - \frac{1}{6} x\delta^3 + O\left(\frac{1}{x\delta}\right) = 0 \qquad \ldots \quad (16)$$

from which δ may be determined by successive approximations leading to the result

$$\delta = \sqrt{\frac{1}{x} \log \frac{x}{2\pi}} + \frac{1}{6x} \log \frac{x}{2\pi} - \frac{1}{2x} + o\left(\frac{1}{x}\right).$$

From this we immediately get Theorem 1.

From $M \to \infty$ we may disregard m/M in (3). In the following we shall therefore first solve the equation

$$m = xB(c, x). \qquad \ldots \quad (17)$$

From this solution we may then find the solution, m^*, say, to (3) as

$$m^* = \frac{m}{1 + m/M} = \frac{Mm}{M+m}. \qquad \ldots \quad (18)$$

(It should be noted that m in (17) is Dodge and Romig's y).

To evaluate $B(c, x)$ we use the normal approximation to the Poisson distribution together with Mill's ratio which give

$$B(c, x) \sim \Phi(\delta\sqrt{x}) \sim 1 - \phi(\delta\sqrt{x})/\delta\sqrt{x} \sim 1 - \frac{1}{\sqrt{x \log \frac{x}{2\pi}}}. \qquad \ldots \quad (19)$$

According to (17) this should be equal to m/x. A comparison with the values tabulated by Dodge and Romig shows that a considerably better approximation may be obtained

by adding a further term equal to $1.9/x$. (It should be noted that this term has not been derived from an asymptotic expansion). From (17) we then get

$$m \sim x - \sqrt{\frac{x}{\log \frac{x}{2\pi}}} + 1.9. \qquad \ldots \quad (20)$$

By inversion it is found to the same degree of approximation that

$$x \sim m + \sqrt{\frac{m}{\log \frac{m}{2\pi}}} - 1.9. \qquad \ldots \quad (21)$$

Combining (7) and (21) we find the following theorem.

Theorem 2 : *Asymptotically the acceptance number c may be expressed as the following function of $m = np_L$*

$$c = m + \sqrt{m \log \frac{m}{2\pi}} + \sqrt{\frac{m}{\log \frac{m}{2\pi}}} + \frac{1}{6} \log \frac{m}{2\pi} - 1.9. \qquad \ldots \quad (22)$$

For the LTPD system we found

$$\frac{c}{n} \sim p_t - 1.28\sqrt{\frac{p_t}{n}}, \qquad \ldots \quad (23)$$

i.e. c/n converges to p_t from below, the difference as usual converging to zero as $1/\sqrt{n}$ because a point $(p_t, \ 0.10)$ of the OC-curve for the sampling plan has been fixed.

The result for the AOQL system just proved shows that

$$\frac{c}{n} \sim p_L + \sqrt{\frac{p_L}{n}} \sqrt{\frac{\log np_L}{2\pi}}, \qquad \ldots \quad (24)$$

i.e. c/n converges to p_L from above but the difference converges to zero at a much slower rate than $1/\sqrt{n}$ because fixing the maximum of the AOQ curve is a requirement essentially different from fixing a point of the OC-curve.

Considered as a limit theorem the first few terms of (22) are sufficient. However, since we also want to develop formulas which may be used for finite lots (22) has the drawback that it is not defined for $m < 2\pi$. Preserving the asymptotic properties we have therefore modified (22) to the following which is defined for $m \geqslant 0.5$:

$$c = m + \sqrt{m - 0.5}\left(\frac{v+1}{\sqrt{v}}\right) + 0.5v - 1.9 \qquad \ldots \quad (25)$$

where $v = \log (1 + m/2\pi)$.

Table 1 shows the relation between c and m. Comparing with the corresponding Dodge-Romig table it will be seen that the approximate m according to (25) deviates at most 0.07 from Dodge and Romig's for $1 \leqslant c \leqslant 40$ which is the tabulated range. The conclusion is that (25) gives a rather accurate solution to equations (2) and (17) apart from the case $c = 0$ which has the solution $m = 0.3679$.

TABLE 1. RELATION BETWEEN m AND c ACCORDING TO (25)

c	m	c	m	c	m
		17.0	11.68	34.0	24.94
0.5	0.73	17.5	12.06	34.5	25.34
1.0	0.89	18.0	12.44	35.0	25.74
1.5	1.09	18.5	12.82	35.5	26.14
2.0	1.33	19.0	13.19	36.0	26.54
2.5	1.59	19.5	13.57	36.5	26.95
3.0	1.88	20.0	13.96	37.0	27.35
3.5	2.17	20.5	14.34	37.5	27.75
4.0	2.48	21.0	14.72	38.0	28.16
4.5	2.80	21.5	15.10	38.5	28.56
5.0	3.12	22.0	15.49	39.0	28.97
5.5	3.45	22.5	15.88	39.5	29.37
6.0	3.78	23.0	16.26	40.0	29.78
6.5	4.12	23.5	16.65	40.5	30.19
7.0	4.46	24.0	17.04	41.0	30.59
7.5	4.80	24.5	17.43	41.5	31.00
8.0	5.15	25.0	17.81	42.0	31.41
8.5	5.49	25.5	18.20	42.5	31.82
9.0	5.85	26.0	18.60	43.0	32.23
9.5	6.20	26.5	18.99	43.5	32.64
10.0	6.55	27.0	19.38	44.0	33.04
10.5	6.91	27.5	19.77	44.5	33.45
11.0	7.27	28.0	20.17	45.0	33.86
11.5	7.63	28.5	20.56	45.5	34.28
12.0	7.99	29.0	20.95	46.0	34.69
12.5	8.35	29.5	21.35	46.5	35.10
13.0	8.72	30.0	21.75	47.0	35.51
13.5	9.08	30.5	22.14	47.5	35.92
14.0	9.45	31.0	22.54	48.0	36.33
14.5	9.82	31.5	22.94	48.5	36.75
15.0	10.19	32.0	23.34	49.0	37.16
15.5	10.56	32.5	23.74	49.5	37.57
16.0	10.94	33.0	24.13	50.0	37.99
16.5	11.31	33.5	24.53	50.5	38.40

$$m^* = \frac{Mm}{M+m} = np_L$$

4. Relation between lot size and sample size

From Theorem 2 it follows that c tends to infinity with $m = np_L$. We therefore need an asymptotic expression for the Poisson distribution with parameter $\bar{m} = n\bar{p}$, $\bar{p} < p_L$, for $\bar{m}$ and $c \sim np_L$ both tending to infinity. The required result is given by the following lemma:

Lemma: *For $c > \bar{m}$ we have asymptotically*

$$1-B(c, \bar{m}) = \frac{\bar{m}}{c-\bar{m}} \frac{1}{\sqrt{2\pi c}} e^{-\bar{m}+c-c \log (c/\bar{m})} (1+O((c-\bar{m})^{-1})). \quad \ldots \quad (26)$$

This lemma is a special case of a theorem given by Blackwell and Hodges (1959).

From (4b) we have

$$z = m+(M-m)(1-B(c, rm)) = m+(M-m)f(m), \qquad \dots \quad (27)$$

say, since c is a function of m given by (25). Minimizing z with respect to m leads to the equation

$$dz/dm = 1+(M-m)f'(m)-f(m) = 0$$

from which we find

$$\log (M-m) = -\log f(m)-\log(-d \log f(m)/dm)+\log(1-f(m)). \qquad \dots \quad (28)$$

This relation gives us M as a function of m. In the following we shall derive asymptotic expansions for each of the three terms on the right hand side of (28) discarding all terms which are $o(1)$. This leads to the following theorem.

Theorem 3 : *The asymptotic relation between sample size and lot size is given by*

$$\log M = m(r-\log r-1)-\sqrt{m-0.5}(\sqrt{v}+1/\sqrt{v}) \log r$$
$$+\frac{1}{2} v(1-\log r)+\frac{1}{2} \log m-\log (r-\log r-1+1/m)$$
$$+0.85 \log r-0.26 \log (1-r)+\log \sqrt{2\pi}-0.1+1/2m, \qquad \dots \quad (29)$$

where $v = \log (1+m/2\pi)$.

Proof: For convenience we write in the following

$$\epsilon = \sqrt{\frac{\log (1+m/2\pi)}{m-0.5}} \qquad \dots \quad (30)$$

so that (25) may be written as

$$c = (m-0.5) \left[1+\epsilon+\frac{1}{(m-0.5)\epsilon}+\frac{1}{2}\epsilon^2-\frac{1.4}{m-0.5}+o\left(\frac{1}{m}\right) \right]$$

from which follows

$$\log \frac{c}{m-0.5} = \epsilon+\frac{1}{(m-0.5)\epsilon}-\frac{2.4}{m-0.5}+o\left(\frac{1}{m}\right)$$

and

$$\log \frac{c}{m} = \log \frac{c}{m-0.5}-\frac{0.5}{m-0.5}+o\left(\frac{1}{m}\right)$$

$$= \epsilon+\frac{1}{(m-0.5)\epsilon}-\frac{2.9}{m-0.5}+o\left(\frac{1}{m}\right).$$

Introducing

$$g(m) = \bar{m}-c+c \log \frac{c}{\bar{m}}$$

$$= m \left[r-\frac{c}{m} (1+\log r)+\frac{c}{m} \log \frac{c}{m} \right]$$

we find

$$g(m) = m(r - \log r - 1) - \left[(m-0.5)\epsilon + \frac{1}{\epsilon} \right] \log r$$

$$+\frac{1}{2}(m-0.5)\epsilon^2(1-\log r)+1+1.9\log r+o(1). \quad \ldots \quad (31)$$

Since
$$\frac{\bar{m}}{c-\bar{m}} = \frac{r}{\frac{c}{m}-r} = \frac{r}{1-r}(1+O(\epsilon))$$

and
$$\sqrt{2\pi c} = \sqrt{2\pi m}(1+O(\epsilon))$$

we finally get from (26) that

$$f(m) = \frac{r}{1-r}\frac{1}{\sqrt{2\pi m}} e^{-g(m)}(1+O(\epsilon)) \quad \ldots \quad (32)$$

from which follows that

$$-\log f(m) = g(m) + \frac{1}{2}\log m + \log\sqrt{2\pi} - \log\frac{r}{1-r} + o(1), \quad \ldots \quad (33)$$

$$-d\log f(m)/dm = (r - \log r - 1) + o(1) \quad \ldots \quad (34)$$

and
$$\log(1-f(m)) = -f(m) + o(f(m)) = o(1).$$

Inserting these results into (28) we find

$$\log(M-m) = m(r-\log r-1) - \left[(m-0.5)\epsilon + \frac{1}{\epsilon}\right]\log r$$

$$+\frac{1}{2}(1-\log r)(m-0.5)\epsilon^2 + \frac{1}{2}\log m$$

$$-\log(r-\log r-1)+1+0.9\log r+\log(1-r)+\log\sqrt{2\pi} \quad \ldots \quad (35)$$

which is identical to (29) apart from the last (constant) term. Numerical investigations have proved that a better approximation is obtained for small m by changing the last term as indicated in (29) and adding a further term equal to $1/2m$.

As explained in Section 2 we may use formulas (18), (25), and (29) to obtain approximations to the Dodge-Romig plans by putting c equal to 0.5, 1.5, ..., solving (25) for m by means of Table 1, and finding *intervals* for M (corresponding to every integer value of c) from (29).

Suppose we want to determine the interval for M corresponding to $c = 6$ and $r = 0.9$. From Table 1 we read $m_1 = 3.45$ for $c = 5.5$ and $m_2 = 4.12$ for $c = 6.5$. Inserting these values into (29) we find $M_1 = 53$ and $M_2 = 72$. Thus, for $53 < M \leqslant 72$ we use $c = 6$ and $np_L = m^* = 3.78M/(M+3.78)$. Numerical results will be discussed in Section 7.

A formal inversion of (29) has been carried out but the result is not useful for practical purposes since the convergence of the series is very slow. A "numerical inversion" is given in Section 7.

Comparing the asymptotic formulas for the two systems of Dodge-Romig plans it will be seen that *an essential difference comes from the second term of the expansions* which is of order $\sqrt{m}$ for the LTPD plans but of order $\sqrt{m \log m}$ for the AOQL plans.

5. The probability of acceptance

Since $c \to np_L$ we shall consider the operating characteristic in the neighbourhood of p_L. By means of the normal approximation to the Poisson distribution we find the probability of acceptance as

$$P(p) \sim B(c, np) \sim \Phi\left(\frac{c-np}{\sqrt{np}}\right) \sim \Phi\left(\frac{c-np}{\sqrt{np_L}}\right).$$

Using (24) we find the value of p giving a specified acceptance probability P as

$$p \sim p_L\left(1+\frac{\sqrt{\log (np_L/2\pi)}-u_P}{\sqrt{np_L}}\right) \qquad \ldots \quad (36)$$

where u_P denotes the P fractile of the normal distribution. It follows that the OC-curve tends to become vertical through $p = p_L$ but that the convergence to the limiting form is very slow.

From (28) and (34) we obtain the producers risk

$$1-P(\bar{p}) = f(m) \sim \frac{1}{M(r-\log r-1)}, \qquad \ldots \quad (37)$$

i.e. *asymptotically the producers risk decreases inversely proportional to lot size.*

This result is the same as for the LTPD plans. It should be noted, however, that (37) gives a rather poor approximation to the producers risk unless M is very large and r is small.

As stated by Hald (1962) without proof there exists a simple asymptotic relation between p_t and p_L for the LTPD plans showing that $p_L \to p_t$. The proof depends on Theorem 2. Similarly, it follows for the AOQL plans that $p_t \to p_L$, p_t being defined as the quality having an acceptance probability of 0.10. From (36) we get for $P = 0.10$ and $m_t = np_t$

$$m_t \sim m+\sqrt{m \log \frac{m}{2\pi}}+1.28\sqrt{m}.$$

Since this formula cannot be used for $m < 2\pi$ it has been taken as model for an improved formula in which $m/2\pi$ has been replaced by $1+m/2\pi$. Retaining the coefficients of the three terms above, changing to logarithms with base 10 and adding a fourth term we find the following rather accurate formula

$$m_t = m+1.52\sqrt{m\log_{10}\left(1+\frac{m}{2\pi}\right)}+1.28\sqrt{m}+6.22\sqrt{\log_{10}\left(1+\frac{m}{2\pi}\right)}$$

where $m = np_L$.

6. Minimum amount of inspection per lot

From (27), (28), and (34) we find

$$z = I(\bar{p})p_L \sim m+1/(r-\log r-1)$$

or

$$I(\bar{p})-n \sim \frac{1}{p_L\log\frac{p_L}{\bar{p}}-(p_L-\bar{p})}, \qquad \ldots \quad (38)$$

i.e. *the average amount of inspection per lot of process average quality apart from sampling inspection is asymptotically independent of lot size.*

This result is also similar to the one found for the LTPD plans. It means that the relative amount of inspection of "rejected" lots of process average quality as compared to sampling inspection tends to zero as $1/\log M$.

7. Numerical investigations

The relation (29) between M and m has been drawn on semi-logarithmic paper for $r = 0.1, 0.2, \ldots, 0.9$, and an inversion of (29) has then been obtained by fitting curves of the form

$$m = \gamma_1\log M+\gamma_2\sqrt{\log M}+\gamma_3\log\ \log M+\gamma_4 \qquad \ldots \quad (39)$$

using suitably selected points. The coefficients found have been expressed as functions of r in the following way :

$$\begin{aligned}
\log\gamma_1 &= 2.3220r + 0.0280/(1-r)-0.0897\\
\log(-\gamma_2) &= 2.5660r+0.0306/(1-r)+0.3645\\
\log\gamma_3 &= 3.0517r+0.0325/(1-r)+0.9158/(1+r)-0.7090\\
\log\gamma_4 &= 2.7291r+0.0307/(1-r)+0.3052/(1+r)-0.0412.
\end{aligned}$$

Table 2 contains a tabulation of these coefficients together with the coefficients in formula (29) after transforming all logarithms to base 10.

The accuracy of (39) as an approximation to (29) depends on r. For $N < 100.000$ and $0.10 \leqslant r \leqslant 0.70$ the error in m will be less than 0.20, for $r = 0.80$ less than 0.30, and for $r = 0.90$ usually less than 0.40 although occasionally as high as 0.70. For most values of M, however, the error will be considerably less than the figures stated.

The asymptotic formulas should not be used for small values of M, i.e. M less than about 15. For such values the exact solution has been given in Table 3.

TABLE 2. COEFFICIENTS β_1 TO β_8 FOR COMPUTING $\log M = \log Np_L$ FROM m AND $r = \bar{p}/p_L$ ACCORDING TO (29) AND COEFFICIENTS γ_1 TO γ_4 FOR COMPUTING m FROM M AND r ACCORDING TO (39). ($\log = \log_{10}$ AND $ln = \log_e$)

	coefficient of \ r	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
β_1	m	0.6091	0.3515	0.2189	0.1374	0.0839	0.0481	0.0246	0.0101	0.0023
β_2	$\sqrt{(m-1/2)\log(1+m/2\pi)}$	1.5174	1.0606	0.7934	0.6038	0.4568	0.3366	0.2351	0.1471	0.0694
β_3	$\sqrt{(m-1/2)/\log(1+m/2\pi)}$	0.6590	0.4606	0.3446	0.2622	0.1984	0.1462	0.1021	0.0639	0.0302
β_4	$\log(1+m/2\pi)$	1.6513	1.3047	1.1020	0.9581	0.8466	0.7554	0.6783	0.6116	0.5527
β_5	$\log m$	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
β_6	$\log(r-\ln r-1+1/m)$	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
β_7	1	-0.4824	-0.2133	-0.0485	0.0751	0.1781	0.2706	0.3599	0.4550	0.5768
β_8	$1/m$	0.2171	0.2171	0.2171	0.2171	0.2171	0.2171	0.2171	0.2171	0.2171
γ_1	$\log M$	1.491	2.569	4.435	7.687	13.406	23.632	42.563	80.891	190.590
γ_2	$\sqrt{\log M}$	-4.520	-8.240	-15.067	-27.663	-51.133	-95.631	-183.105	-371.792	-954.773
γ_3	$\log \log M$	2.916	5.071	9.066	16.595	31.073	59.655	118.639	253.302	699.085
γ_4	1	3.494	6.271	11.380	20.874	38.745	73.074	141.590	291.834	763.170

TABLE 3. TABLE OF c AS FUNCTION OF $M = Np_L$ AND $r = \bar{p}/p_L$ FOR $M \leqslant 15$

c \ r	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	1−13	1−7	1−5	1−3	1−3	1−2	1−2	1−2	1−2
1	14−15	8−15	6−15	4−15	4−13	3−10	3−8	3−7	3−6
2					14−15	11−15	9−15	8−15	7−13
3									14−15

For values of (M, r) not given in the table the asymptotic formulas may be used.

For $c = 0$ use the following $m = m(M)$

M	1	2	3	4–5	6–8	9–13
$m(M)$	0.27	0.31	0.33	0.34	0.35	0.36

For $c > 0$ use m^* from Table 1.

It is rather difficult to carry out a numerical comparison of the exact and the asymptotic solution because it is impossible to read the exact solution with sufficient accuracy from Dodge and Romig's graphs and their tables give sampling plans for a rather coarse lot size grouping. Table 4 contains, however, some characteristic results of the comparisons which has been carried out for tabulated Dodge-Romig plans corresponding to upper end-points of lot size intervals. The table shows for each M

the tabulated Dodge-Romig plan, the "asymptotic plan", the average amount of inspection, and the AOQL. It will be seen that there is no essential difference between the two sets of plans.

TABLE 4. COMPARISON OF "ASYMPTOTIC PLANS" ACCORDING TO (39) AND TABULATED DODGE-ROMIG PLANS

$100p_L$	r	M	Dodge and Romig c	m	asympt. c	m^*	$I^D(\bar{p})p_L$	$I^A(\bar{p})p_L$	$100p_L^D$	$100p_L^A$
0.50	0.10	50	1	0.83	1	0.87	0.99	1.05	0.50	0.47
		500	2	1.38	2	1.33	1.58	1.51	0.50	0.51
	0.50	20	2	1.28	2	1.25	1.79	1.73	0.50	0.51
		50	3	1.88	3	1.81	2.63	2.47	0.50	0.52
		500	6	3.78	7	4.42	5.44	5.43	0.50	0.50
	0.90	20	3	1.78	3	1.72	3.22	3.03	0.50	0.52
		50	5	2.98	5	2.94	5.58	5.41	0.50	0.51
		500	16	10.7	15	10.0	20.4	20.8	0.50	0.50
2.0	0.10	80	1	0.84	1	0.88	1.10	1.17	2.0	1.9
		200	2	1.40	2	1.32	1.48	1.39	1.9	2.1
		2000	3	1.90	4	1.88	1.99	1.97	2.0	2.1
	0.50	10	1	0.78	1	0.82	1.32	1.41	2.0	1.9
		20	2	1.30	2	1.24	1.83	1.71	2.0	2.1
		80	3	1.90	4	2.40	3.16	3.00	2.0	2.1
		200	5	3.10	5	3.08	4.12	4.07	2.0	2.0
		2000	9	5.80	10	6.52	7.51	7.68	2.0	2.0
	0.90	20	3	1.80	3	1.72	3.29	3.03	2.0	2.1
		80	7	4.20	6	3.60	7.17	7.20	2.0	2.0
		200	11	7.00	10	6.34	12.4	12.5	2.0	2.0
		2000	27	19.1	28	20.0	39.8	40.4	2.0	2.0
10.0	0.10	50	1	0.8	1	0.9	0.95	1.09	10.3	9.2
		400	2	1.4	2	1.3	1.56	1.43	9.8	10.5
		10000	3	1.9	4	2.5	2.37	2.57	10.2	10.2
	0.50	50	3	1.9	3	1.8	2.67	2.45	9.8	10.4
		100	4	2.5	4	2.4	3.39	3.16	9.9	10.3
		400	6	3.8	6	3.7	5.17	4.88	9.9	10.2
		1000	8	5.0	8	5.1	6.13	6.41	10.2	10.0
		10000	12	8.0	13	8.7	10.7	10.5	9.9	10.0
	0.90	20	3	1.8	3	1.7	3.28	2.97	9.8	10.5
		50	5	3.0	5	2.9	5.67	5.24	9.9	10.3
		100	8	4.9	7	4.3	8.35	8.49	10.0	10.0
		400	15	10.0	14	9.2	18.6	18.0	9.9	10.0
		1000	23	16.0	21	14.5	28.5	28.9	10.0	10.0

Figure 1 shows the relation between M and m according to (39) for $r = 0.1, 0.2, \dots, 0.9$. The relation (25) between m and c has been given on the vertical scale. From given values of $M = Np_L$ and $r = \bar{p}/p_L$ we may thus read (m, c) on Fig. 1 and then compute m^* from (18). To make it easy for the reader to draw his own diagram the necessary values of (M, m) have been recorded in Table 5.

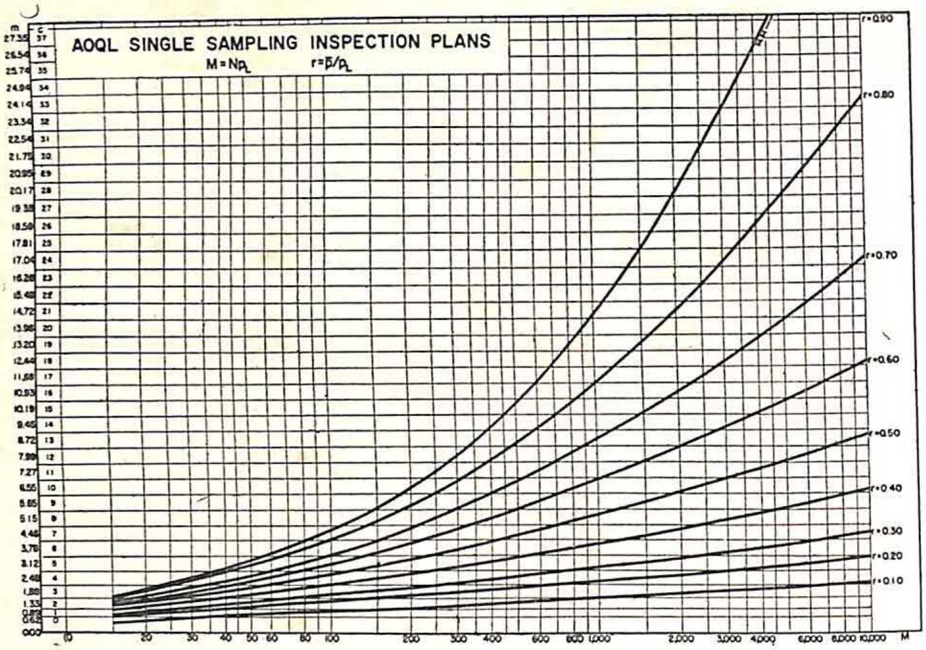

Fig. 1. Relation between M, m, and c according to formulas (39) and (25).

TABLE 5. TABLE OF m AS FUNCTION OF $M = Np_L$ AND $r = \bar{p}/p_L$ ACCORDING TO FORMULA (39)

M	r=0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
15	0.55	0.71	0.89	1.08	1.25	1.36	1.43	1.61	1.14
20	0.61	0.79	1.00	1.22	1.41	1.56	1.67	1.95	1.99
40	0.76	1.00	1.27	1.57	1.86	2.10	2.30	2.69	3.11
100	0.96	1.28	1.67	2.12	2.60	3.05	3.48	4.07	4.54
200	1.12	1.52	2.01	2.61	3.27	3.98	4.71	5.67	6.43
400	1.29	1.77	2.38	3.15	4.05	5.08	6.25	7.78	9.30
1000	1.53	2.13	2.91	3.94	5.22	6.79	8.74	11.40	14.77
2000	1.72	2.41	3.35	4.60	6.21	8.27	10.95	14.73	20.19
4000	1.91	2.71	3.81	5.30	7.28	9.90	13.42	18.56	26.69
10000	2.17	3.12	4.44	6.29	8.81	12.26	17.06	24.32	30.88

ACKNOWLEDGEMENT

The basic theoretical results of this paper were developed by A. Hald during a stay at the Statistical Techniques Research Group, Princeton University, and my thanks are due to members of the staff and in particular to Dr. V. S. Varadarajan for helpful discussions. Our thanks are also due to Mr. K. West Andersen for carrying out some of the computations.

REFERENCES

BLACKWELL, D. and HODGES, J. L. (1959): The probability in the extreme tail of a convolution. *Ann. Math. Stat.*, **30**, 1113–1120.

DODGE, H. F. and ROMIG, H. G. (1941): Single sampling and double sampling inspection tables. *Bell. System Technical Journal*, **20**, 1–61, also reprinted in H.F. Dodge and H. G. Romig : Sampling Inspection Tables, John Wiley, New York, 2-ed. 1959.

HALD, A. (1962): Some limit theorems for the Dodge-Romig LTPD single sampling inspection plans. *Technometrics*, **4**, 497–513.

Paper received : June, 1962.
Revised : January, 1963.

LONG-CHAIN POLYMERS AND SELF-AVOIDING RANDOM WALKS II

By J. M. HAMMERSLEY

University of Oxford

SUMMARY. This continuation of a previous paper discusses the dependence of the connective constant upon dimensionality.

Hammersley (1963) gave the background and notation for self-avoiding Pólya walks. This paper is a direct continuation with equations and theorems numbered serially (so that references to equations (1) to (62) are in fact references to this previous paper).

We define

$$\lambda(d) = e^{\kappa}, \qquad \dots \quad (63)$$

where κ is the connective constant introduced in (6); and we study the dependence of λ upon d.

To emphasize our concern with d, we write $F_n^{(d)}$ for the class of all self-avoiding n-stepped Pólya walks on a d-dimensional lattice, starting from the origin. We write $f_n^{(d)}$ for the number of distinct members of $F_n^{(d)}$; and we define $f_0^{(d)} = 1$. We introduce the generating function

$$\varphi^{(d)}(x) = \sum_{n=0}^{\infty} f_n^{(d)} x^n, \quad |x| < 1/\lambda(d). \qquad \dots \quad (64)$$

In fact, it follows from (6) and (63) that $1/\lambda(d)$ is the radius of convergence of $\varphi^{(d)}(x)$.

Let δ be an integer satisfying $1 \leqslant \delta < d$. Any member of $F_n^{(d)}$ can be uniquely represented as an ordered sequence of n unit vectors $V_1 V_2 \dots V_n$ in d-dimensional space, where V_i is the vector joining the $(i-1)$-th to the i-th point of the walk. Some of these vectors V_i may lie in the subspace spanned by the first δ coordinate axes. Let $V_{i_1} V_{i_2} \dots V_{i_m}$ be the subsequence of all vectors with this property. If and only if $V_{i_1} V_{i_2} \dots V_{i_m}$ is a member of $F_m^{(\delta)}$, we say that the original walk $V_1 V_2 \dots V_n$ is a member of $F_n^{(d,\delta)}$. This is a definition of the class $F_n^{(d,\delta)}$. In this definition we adopt the convention that $F_0^{(\delta)}$ is the empty set of vectors; and consequently a walk of $F_n^{(d)}$ belongs to $F_n^{(d,\delta)}$ if none of its vectors V_i are parallel to any of the first δ coordinate axes. We define

$$\varphi^{(d,\delta)}(x) = \sum_{n=0}^{\infty} f_n^{(d,\delta)} x^n, \qquad \dots \quad (65)$$

with the usual gloss that $f_0^{(d,\delta)} = 1$.

Theorem 11 :

$$\varphi^{(d)}(x) \geqslant \varphi^{(d,\delta)}(x) = \varphi^{(d-\delta)}(x)\varphi^{(\delta)}[x\varphi^{(d-\delta)}(x)]. \qquad \dots \quad (66)$$

Proof of Theorem 11 : Since $F_n^{(d,\delta)}$ is a subclass of $F_n^{(d)}$, we have $f_n^{(d,\delta)} \leqslant f_n^{(d)}$; and accordingly the first half of (66) is trivial.

Suppose, then, that $V_1 V_2 \dots V_n$ is a sequence of vectors representing a typical member of $F_n^{(d,\delta)}$; and that $V_{i_1} V_{i_2} \dots V_{i_m}$ (possibly empty) is the subsequence representing the derived member of $F_m^{(\delta)}$. Then the interspersed sequences

$$V_1 V_2 \dots V_{i_1-1};\; V_{i_1+1} V_{i_1+2} \dots V_{i_2-1};\; \dots;\; V_{i_m+1} V_{i_m+2} \dots V_n \qquad \dots \quad (67)$$

consist of vectors in the space spanned by the last $d-\delta$ coordinate axes, and are consequently members of

$$F_{i_1-1}^{(d-\delta)},\quad F_{i_2-i_1-1}^{(d-\delta)}, \dots, F_{n-i_m}^{(d-\delta)} \qquad \dots \quad (68)$$

respectively. Some or all of the sequences in (67) may be empty (e.g. whenever $i_{j+1}-i_j-1 = 0$) and the corresponding entry in (68) will then be $F_0^{(d-\delta)}$. Conversely if we take arbitrary members of the classes (68) and use their representative sequences to intersperse the vectors of an arbitrary member of $F_m^{(\delta)}$, we shall obtain a member of $F_n^{(d,\delta)}$; and distinct arbitrary choices lead to distinct members of $F_n^{(d,\delta)}$. Hence

$$f_n^{(d,\delta)} = \Sigma f_m^{(\delta)} f_{i_1-1}^{(d-\delta)} f_{i_2-i_1-1}^{(d-\delta)} \dots f_{n-i_m}^{(d-\delta)}, \qquad \dots \quad (69)$$

where the summation is over all integers i_j such that

$$1 \leqslant i_1 < i_2 < \dots < i_m \leqslant n \qquad \dots \quad (70)$$

and over all m satisfying $0 \leqslant m \leqslant n$. (When $m = 0$, the term in (69) is to be taken as $f_n^{(d-\delta)}$.) However, if we expand the right-hand side of (66) as a power series in x, using (64), we find that the coefficient of x^n is precisely the right-hand side of (69). This completes the proof of Theorem 11.

Theorem 12 : *For all positive integers a and b*

$$\lambda(a+b) \geqslant \lambda(a)+\lambda(b). \qquad \dots \quad (71)$$

Proof of Theorem 12 : Writing $\delta = a$ and $d-\delta = b$ in (66) we have

$$\varphi^{(a+b)}(x) \geqslant \varphi^{(b)}(x)\varphi^{(a)}[x\varphi^{(b)}(x)]. \qquad \dots \quad (72)$$

From the left-hand side of (6), we have for any d

$$\varphi^{(d)}(x) = \sum_{n=0}^{\infty} f_n^{(d)} x^n \geqslant \sum_{n=0}^{\infty} (e^{\kappa}x)^n = [1-\lambda(d)x]^{-1} \qquad \dots \quad (73)$$

by virtue of (63). Thus

$$\begin{aligned} \varphi^{(a+b)}(x) &\geqslant \varphi^{(b)}(x)[1-\lambda(a)x\varphi^{(b)}(x)]^{-1} \\ &\geqslant [1-\lambda(b)x]^{-1}\{1-\lambda(a)x/[1-x\lambda(b)]\}^{-1} \\ &= \{1-[\lambda(a)+\lambda(b)]x\}^{-1}. \end{aligned} \qquad \dots \quad (74)$$

However, the radius of convergence of the right-hand side of (74) is $[\lambda(a)+\lambda(b)]^{-1}$; and this must be at least as great as the radius of convergence of the left-hand side of (74), namely $1/\lambda(a+b)$. This proves (71).

Theorem 13 : *The function $\psi_a(x)$, defined by*

$$\psi_a(x) = \begin{cases} x/\varphi^{(a)}(1/x), & x > \lambda(a) \\ 0, & x = \lambda(a), \end{cases} \qquad \dots \quad (75)$$

satisfies the functional inequality

$$\psi_{a+b}(x) \leqslant \psi_a[\psi_b(x)], \quad x \geqslant \lambda(a+b), \qquad \dots \quad (76)$$

where a and b are positive integers. The derivative of $\psi_a(x)$ exists and satisfies

$$0 < \psi_a'((x) < 1, \quad x > \lambda(a). \qquad \dots \quad (77)$$

Proof of Theorem 13 : By substituting (75) into (72), we obtain (76). To prove (77), we first note that the coefficients in the expansion of $\varphi^{(a)}(x)$ are all

positive integers; so that $\varphi^{(a)}(x)$ is a positive strictly increasing function for $0 \leqslant x < 1/\lambda(a)$. Thus $\psi_a(x)$ is also a positive strictly increasing function for $x > \lambda(a)$. Within its circle of convergence $\varphi^{(a)}(x)$ is analytic; and hence $\psi_a(x)$ is differentiable for $x > \lambda(a)$. Finally

$$0 < \psi_a'(x) = \frac{1}{\varphi^{(a)}(1/x)}\left\{1-\frac{\varphi^{(a)\prime}(1/x)}{x\varphi^{(a)}(1/x)}\right\} < \frac{1}{\varphi^{(a)}(1/x)} \leqslant \frac{1}{f_0^{(a)}} = 1, \qquad \dots \quad (78)$$

which establishes (77).

Theorem 14 : *For any pair of positive integers a and d,*

$$\int_{\lambda(a)}^{\lambda(ad)} \frac{dx}{x-\psi_a(x)} > d-1. \qquad \dots \quad (79)$$

In the particular case when $a = 1$, the inequality (79) *yields*

$$\lambda(d) > (2d-1)-\log(2d-1), \qquad \dots \quad (80)$$

which establishes (5).

Proof of Theorem 14 : From (73), we see that $\varphi^{(a)}(x) \to \infty$ as $x \to 1/\lambda(a)$ from below. Hence, by (75), $\psi_a(x)$ is continuous at $x = \lambda(a)$. By integrating (78) from $x = \lambda(a)$ to $x = y$, we get

$$\psi_a(y)-\psi_a(\lambda(a)) \leqslant y-\lambda(a) ; \qquad \dots \quad (81)$$

and hence, by (75),

$$y-\psi_a(y) \geqslant \lambda(a) > 0. \qquad \dots \quad (82)$$

Therefore the integral

$$I(x) = \int_{\lambda(a)}^{x} \frac{dy}{y-\psi_a(y)}, \quad x \geqslant \lambda(a), \qquad \dots \quad (83)$$

has a bounded continuous positive integrand; and therefore $I(x)$ has a positive derivative and is a strictly increasing function of x. Since $\psi_a(y) \geqslant 0$, we have

$$I(x) \geqslant \int_{\lambda(a)}^{x} \frac{dy}{y} \to \infty \text{ as } x \to \infty; \qquad \dots \quad (84)$$

so $I(x)$ increases strictly from 0 to ∞ as x goes from $\lambda(a)$ to ∞. Consequently there exists a uniquely defined inverse function $\chi(\xi)$ such that

$$\xi = \int_{\lambda(a)}^{\chi(\xi)} \frac{dy}{y-\psi_a(y)}, \quad \xi \geqslant 0. \qquad \dots \quad (85)$$

Since $I(x)$ has a positive derivative, $\chi(\xi)$ must be differentiable and have a positive derivative; and differentiation of (85) yields

$$\chi'(\xi) = \chi(\xi)-\psi_a[\chi(\xi)]. \qquad \dots \quad (86)$$

From (84), $\chi(\xi) \geqslant \lambda(a)$. Hence the right-hand side of (86) is differentiable in view of Theorem 13 and the existence of $\chi'(\xi)$. Thus the left-hand side of (86) is differentiable. Consequently $\chi(\xi)$ is twice differentiable; and (86) yields

$$\chi''(\xi) = \chi'(\xi)\{1-\psi_a'[\chi(\xi)]\}, \qquad \dots \quad (87)$$

and this is strictly positive by virtue of (77). The mean-value theorem now gives, for any $\xi \geqslant 1$ and some θ satisfying $0 < \theta < 1$,

$$\chi(\xi-1) = \chi(\xi)-\chi'(\xi)+\tfrac{1}{2}\chi''(\xi-\theta) > \chi(\xi)-\chi'(\xi) = \psi_a[\chi(\xi)], \qquad \dots \quad (88)$$

where the last step results from (86).

For the sake of a contradiction, let us now suppose that

$$\chi(d-1) \geqslant \lambda(ad). \qquad \dots \quad (89)$$

We then have, by repeated use of (76) and (88) alternately,

$$\begin{aligned} 0 = \psi_{ad}[\lambda(ad)] &\leqslant \psi_{ad}[\chi(d-1)] \leqslant \psi_{a(d-1)}[\psi_a[\chi(d-1)]] \\ &< \psi_{a(d-1)}[\chi(d-2)] \leqslant \psi_{a(d-2)}[\psi_a[\chi(d-2)]] \\ &< \psi_{a(d-2)}[\chi(d-3)] \leqslant \ldots < \psi_a[\chi(0)] = \psi_a[\lambda(a)] = 0. \qquad \ldots (90) \end{aligned}$$

This contradiction shows that (89) is false. Hence

$$\chi(d-1) < \lambda(ad); \qquad \ldots (91)$$

and therefore, from (85),

$$d-1 = \int_{\lambda(a)}^{\chi(d-1)} \frac{dy}{y-\psi_a(y)} < \int_{\lambda(a)}^{\lambda(ad)} \frac{dy}{y-\psi_a(y)}. \qquad \ldots (92)$$

This proves (79). By substituting (64) and (75) into (79), we find that (79) is equivalent to

$$\int_{\lambda(a)}^{\lambda(ad)} \left\{1+\left[\sum_{n=1}^{\infty} \frac{f_n^{(a)}}{x^n}\right]^{-1}\right\} \frac{dx}{x} > d-1. \qquad \ldots (93)$$

Let us now approximate the integral in (93). We have

$$f_1^{(a)} = 2a; \qquad f_2^{(a)} = 2a(2a-1) \qquad \ldots (94)$$

and

$$2a(2a-1)^2 \leqslant f_n^{(a)} \leqslant 2a(2a-1)^{n-1}, \qquad n \geqslant 3. \qquad (95)$$

Hence

$$\frac{2a}{x}+\frac{2a(2a-1)}{x^2}+\frac{2a(2a-1)^2}{x^2(x-1)} \leqslant \sum_{n=1}^{\infty} \frac{f_n^{(a)}}{x^n} \leqslant \frac{2a}{x}+\frac{2a(2a-1)}{x(x-2a+1)}. \qquad \ldots (96)$$

A little manipulation of (96) leads to

$$\frac{x+1}{2a}+\frac{(a-1)(2a-1)^2}{a[x^2+2(a-1)x+2(a-1)(2a-1)]} \geqslant 1+\left[\sum_{n=1}^{\infty} \frac{f_n^{(a)}}{x^n}\right]^{-1} \geqslant \frac{x+1}{2a}. \qquad \ldots (97)$$

Since a walk whose steps are always in the positive directions of the coordinate axes is necessarily self-avoiding, we have $f_n^{(a)} \geqslant a^n$ with the consequence that $\lambda(a) \geqslant a$. Hence, by (97),

$$\begin{aligned} 0 &\leqslant \int_{\lambda(a)}^{\lambda(ad)} \left\{1+\left[\sum_{n=1}^{\infty} \frac{f_n^{(a)}}{x^n}\right]^{-1}\right\} \frac{dx}{x} - \left\{\frac{\lambda(ad)+\log\lambda(ad)-\lambda(a)-\log\lambda(a)}{2a}\right\} \\ &\leqslant \int_{\lambda(a)}^{\lambda(ad)} \frac{(a-1)(2a-1)^2 dx}{ax[x^2+2(a-1)x+2(a-1)(2a-1)]} \leqslant 4a(a-1)\int_a^{\infty} \frac{dx}{x^3} = 2\left(1-\frac{1}{a}\right). \qquad \ldots (98) \end{aligned}$$

Thus (93) yields

$$\lambda(ad)+\log\lambda(ad) > \lambda(a)+\log\lambda(a)+2ad-2a-4(a-1). \qquad \ldots (99)$$

When $a = 1$, we have $\lambda(1) = 1$; and (99) gives

$$\lambda(d)+\log\lambda(d) > 2d-1 \geqslant [(2d-1)-\log(2d-1)]+\log[(2d-1)-\log(2d-1)]. \ \ldots (100)$$

The left-hand side of (100) is an increasing function of $\lambda(d)$; and (80) follows.

It is somewhat disappointing that the case $a > 1$ does not lead to a stronger version of (80). It follows from (98) that (93) can never yield anything stronger than

$$\lambda(ad)+\log\lambda(ad) > \lambda(a)+\log\lambda(a)-2ad-2a > 2ad-1, \qquad \ldots (101)$$

on using (80) on the right-hand side of (101); and, of course, the left-hand side of (101) may not be true. If indeed (80) can be strengthened, then presumably inequalities such as (66) and (88) are capable of sharpening.

REFERENCE

HAMMERSLEY, J. M. (1963) : Long-chain polymers and self-avoiding walks. *Sankhyā*, Series A, **25**, 25–34.

Paper received : March, 1963.

DENSITY IN THE LIGHT OF PROBABILITY THEORY-III

By E. M. PAUL

Indian Statistical Institute

SUMMARY. Using magnification methods, we prove the Erdös-Wintner theorem on additive arithmetical functions having distributions, and obtain some generalizations of it. It is also shown that if each of a finite collection of additive functions has a distribution, they have a joint distribution in the sense of logarithmic density.

1. Notation and terminology

In this paper, we shall use the specific space X and the measure P in it that were used in the previous two papers (Paul, 1962a and 1962b), in connection with logarithmic density. $q_1 = 2, q_2, \ldots$ will stand for the prime numbers. If A is any set of real numbers, $\mu(A)$ will stand for the sum of the reciprocals of the positive integers in A. For each positive integer n, A_n will stand for the set of positive integers of the form $q_1^{x_1} \ldots q_n^{x_n}$, the exponents being non-negative integers. If S is any set of positive integers,

$$\lim_n \inf \frac{\mu(S \cap A_n)}{\mu(A_n)} \text{ and } \lim_n \sup \frac{\mu(S \cap A_n)}{\mu(A_n)}$$

will be called respectively the lower and upper Π-densities of S. $\mu(A_n)$ is of course equal to $\left(1-\frac{1}{q_1}\right)^{-1}\left(1-\frac{1}{q_2}\right)^{-1} \ldots \left(1-\frac{1}{q_n}\right)^{-1}$. If S is any set of positive integers,

$$\underline{\lim}_{\frac{\beta}{\alpha}\to\infty} \frac{\mu\{(\alpha, \beta] \cap S\}}{\mu(\alpha, \beta]} \text{ and } \overline{\lim}_{\frac{\beta}{\alpha}\to\infty} \frac{\mu\{(\alpha, \beta] \cap S\}}{\mu(\alpha, \beta]}$$

are called respectively the lower and upper strong logarithmic densities of S; they will be denoted by $\bar{\lambda}^*(S)$ and $\underline{\lambda}^*(S)$. In this paper $\bar{\delta}$, $\underline{\delta}$, $\bar{\lambda}$ and $\underline{\lambda}$ stand respectively for upper natural, lower natural, upper logarithmic and lower logarithmic densities. It is well known (Tsuji, p. 121, for example) that for every set S,

$$\underline{\delta} \leqslant \underline{\lambda}^* \leqslant \underline{\lambda} \leqslant \bar{\lambda} \leqslant \bar{\lambda}^* \leqslant \bar{\delta}.$$

Now let J consisting of $j_1 < j_2 < \ldots$ be a sequence of positive integers. A set S of positive integers will be said to be right-complete with respect to J in case for every positive integer r and arbitrary m such that $j_{r-1} < m \leqslant j_r$, $q_1^{x_1} \ldots q_m^{x_m} \in S$ implies that $q_1^{x_1} \ldots q_m^{x_m} q_{(j_r+1)}^{x_{(j_r+1)}} \ldots q_k^{x_k} \in S$; here the x's are arbitrary non-negative integers and k is any integer $> j_r$. This definition is in accordance with that in the abstract theory that was presented in the first paper (Paul, 1962a). Instead of the abstract space X_1, we now have the set of integers of the form $q_1^{x_1} q_2^{x_2} \ldots q_{j_1}^{x_{j_1}}$; in place of

X_2 we now have the set of integers of the form $q_{(j_1+1)}^{x_{(j_1+1)}} \dots q_{j_2}^{x_{j_2}}$, and so on; the exponents are arbitrary non-negative integers. In each space x_m, the 0-element will be the number $1 = q_{j_{(m-1)}+1}^{0} \dots q_{j_m}^{0}$. We can now talk of the basic vectors of a set of positive integers that is right-complete with respect to J. It will be recalled that we adopted the same procedure in connection with the generalized magnification theorem (Paul, 1962b), $M_J^U(S)$ and $M_J^L(S)$ have been defined there, for an arbitrary set S.

Now let S be right-complete with respect to J. A basic vector $(x_1, \dots, x_m, 0, 0, \dots)$, with x_m as the last positive coordinate, will be said to arise at the r-th stage in case $j_{(r-1)}+1 \leqslant m \leqslant j_r$. Any number of the form $q_1^{x_1} \dots q_m^{x_m} q_{(j_r+1)}^{x_{(j_r+1)}} \dots q_k^{x_k}$ where $x_{(j_r+1)}, \dots, x_k$ are arbitrary non-negative integers and $k \geqslant (j_r+1)$ is a member of S; it will be said to arise at the r-th stage. We shall say that the number $q_1^{x_1} \dots q_k^{x_k}$ enters S through the basic vector $(x_1, \dots, x_m, 0, 0, \dots)$. The set of members of S that arise at the r-th stage will be denoted by S_r. We shall denote $M_J(S_r)$ by D_r.

2. Some convergence theorems

Lemma 1 : *If J is arbitrary and S is a set of positive integers right-complete with respect to J, $P\{M_J(S)\} = \underline{\lambda}(S)$.*

Proof : Let us put $\lambda = \underline{\lambda}(S)$ and $\tau = P\{M_j(S)\}$. Evidently $\lambda \geqslant \tau$.

Take any $\epsilon > 0$. Take a positive integer r such that

$$\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right) \cdots \left(1-\frac{1}{q_{j_s}}\right) > \frac{1}{2 \log q_{j_s}} \text{ for all } s \geqslant r \qquad \dots \quad (2)$$

and

$$P(D_1 U \dots U D_r) > \tau - \epsilon. \qquad \dots \quad (3)$$

For (2), we use Merten's theorem.

Now put

$$\lambda' = \lim_{t\to\infty} \frac{\mu\{(q_{j_r}, q_{j_t}] \cap S\}}{\log q_{j_t}}$$

$$= \lim_{t\to\infty} \frac{\mu\{(q_{j_r}, q_{j_t}] \cap S\}}{\log q_{j_t} - \log q_{j_r}}. \qquad \dots \quad (4)$$

Clearly

$$\lambda' \geqslant \lambda. \qquad \dots \quad (5)$$

Till the end of the proof, L will stand for the interval

$$(q_{j_r}, q_{j_{(r+m)}}].$$

Now choose a positive integer m such that the following four conditions hold.

$$\frac{\mu(L \cap S)}{\mu(L)} \leqslant \lambda' + \epsilon \qquad \dots \quad (6)$$

$$\frac{\mu(L)}{\log q_{j_{(r+m)}}} > 1 - \epsilon \qquad \dots \quad (7)$$

$$\frac{\mu(L \cap S)}{\mu(L)} > \lambda' - \epsilon \qquad \dots \quad (8)$$

$$\frac{\mu\{L \cap (S_1 \cup \dots \cup S_r)\}}{\mu(L)} < T + \epsilon \qquad \dots \quad (9)$$

where $T = P(D_1 U \dots U D_r)$.

There are infinitely many values of m satisfying (6); from among them, we choose an m so large that (7), (8) and (9) hold. For (9), we recall that $(S_1 \cup \dots \cup S_r)$ has natural and hence strong logarithmic density equal to $P(D_1 U \dots U D_r)$.

$$\frac{\mu\{[S - (S_1 \cup \dots \cup S_r)] \cap L\}}{\mu(L)} > \lambda' - T - 2\epsilon, \qquad \dots \quad (10)$$

by (8) and (9).

Again, $$P(D_{r+1} U D_{r+2} U \dots U D_{r+m}) \geqslant (1-\tfrac{1}{2})(1-1/3) \dots \left(1 - \frac{1}{q_{j_{(r+m)}}}\right) \times \mu\{[S-(S_1 \cup \dots \cup S_r)] \cap L\}. \qquad \dots \quad (11)$$

In order to see this, we note that

$$\{S - (S_1 \cup \dots \cup S_r)\} \cap L \subset S_{r+1} \cup \dots \cup S_{r+m} \qquad \dots \quad (12)$$

Now let $(x_1, x_2, \dots, x_{j_{(r+1)}}, 0, 0, \dots)$ be a basic vector arising at the $(r+1)$-th stage. Sum of reciprocals of all positive integers $\leqslant q_{j(r+m)}$ and entering S through this basic vector is

$$\leqslant \frac{1}{2^{x_1} \dots q_{j_{(r+1)}}^{x_{j(r+1)}}} \times \sum \frac{1}{q_{(j_{(r+1)}+1)}^{v_1} \dots q_{j_{(r+m)}}^{v_A}},$$

the v's taking arbitrary non-negative integral values,

$$= \frac{1}{2^{x_1} \dots q_{j_{(r+1)}}^{x_{j(r+1)}}} \times \frac{1}{\left(1 - \frac{1}{q_{(j_{(r+1)}+1)}}\right) \dots \left(1 - \frac{1}{q_{j_{(r+m)}}}\right)}$$

$= \{p$-measure of the cylinder set whose base is the point

$$(x_1, \dots, x_{j_{(r+1)}})\} \times \frac{1}{\left(1 - \frac{1}{2}\right) \dots \left(1 - \frac{1}{q_{j_{(r+m)}}}\right)}.$$

Thus

$$(1-\tfrac{1}{2}) \dots \left(1 - \frac{1}{q_{j_{(r+m)}}}\right) \times \mu\{S_{(r+1)} \cap (0, q_{j_{(r+m)}}]\} \leqslant P(D_{r+1}).$$

Similarly $\qquad (1-\frac{1}{2})\dots\left(1-\dfrac{1}{q_{j_{(r+m)}}}\right)\times\mu\{S_{(r+2)}\bigcap(0,q_{j_{(r+m)}}]\}\leqslant P(D_{r+2}),$

and so on. Adding, we get

$$P(D_{r+1}+\dots+D_{r+m})\geqslant(1-\tfrac{1}{2})\dots\left(1-\frac{1}{q_{j_{(r+m)}}}\right)\times\mu\{(S_{r+1}\bigcup\dots\bigcup S_{r+m})\bigcap L\}.$$

An application of step (12) now gives step (11). From (11) we get

$$P(D_{r+1}\,U\dots U\,D_{r+m})\geqslant(\lambda'-T-2\epsilon).\quad\mu(L),\frac{1}{2\log q_{j_{(r+m)}}}\qquad\dots\quad(13)$$

by (2) and (10).

Now $$(\lambda'-\epsilon)-\tau\leqslant\frac{\mu(S\bigcap L)}{\mu(L)}-P(D_{1U}\dots U\,D_{r+m})\text{ by (8)},$$

$$\leqslant\lambda'+\epsilon-\left\{T+(\lambda'-T-2\epsilon).\;\mu(L).\frac{1}{2\log q_{j_{(r+m)}}}\right\}\text{ by (6) and (13)},$$

$$\leqslant\lambda'+\epsilon-\left\{T+\frac{1}{2\log q_{j_{(r+m)}}}(\lambda'-T-2\epsilon)(1-\epsilon)\log q_{j_{(r+m)}}\right\}\text{ by (7)}$$

$$\leqslant\lambda'+\epsilon-\{T+\tfrac{1}{2}(\lambda'-T-2\epsilon)(1-\epsilon)\}$$

$$\leqslant\lambda'+\epsilon-\{\tau-\epsilon+\tfrac{1}{2}(\lambda'-\tau-2\epsilon)(1-\epsilon)\}\text{ by (3)}.$$

From this inequality, we get by elementary algebraical calculations

$$\lambda'-\tau\leqslant\frac{\epsilon(8-2\epsilon)}{1-\epsilon}\;\therefore\;\lambda-\tau\leqslant\frac{\epsilon(8-2\epsilon)}{1-\epsilon}.$$

Making $\epsilon\to0$, we get $\lambda\leqslant\tau$.

Corollary : *Let J be arbitrary. Let S be a set of positive integers that is right-complete with respect to J. Then*

$$\underline{\lambda}(S)=\underline{\lim_{n}}\frac{\mu\{(0,q_{jn}]\cap S\}}{\log q_{jn}}.$$

Proof: The last expression $=\lambda'$. In (4), λ' appears as dependent on ϵ (through r); but the definition of λ' shows that it is really independent of ϵ. Now we use the inequality

$$\lambda'-\tau\leqslant\frac{(8-2\epsilon)}{1-\epsilon}.$$

Lemma 2 : *If J is arbitrary, and S is any set of positive integers,*

$$P\{M_J^U(S)\}\geqslant\underline{\lambda}(S).$$

Proof: The proof is similar to that of the magnification theorem (see Paul, 1962a, p. 108). As in Paul (1962b), let us call the

$$X_1\dots X_{j_1}\text{-space }Y_1,\,X_{(j_1+1)}\dots X_{j_2}\text{-space }Y_2,$$

and so on. The space X is now regarded as the space $Y_1\, Y_2 \dots$. Let y_1 be the point $(x_1, \dots x_{j_1})$ of Y_1. By $\nu(y_1)$ we shall mean the number $2^{x_1}\, 3^{x_2} \dots q_{j_1}^{x_{j_1}}$. Similarly if y_2 is the point $(x_{j_1+1}, \dots, x_{j_2})$ of y_2, $\nu(y_2)$ will stand for the number $q_{(j_1+1)}^{x_{(j_1+1)}} \dots q_{j_2}^{x_{j_2}}$; and so on.

Let A_k be the set of vectors $(y_1, y_2 \dots)$ such that $y_k \neq (0, 0, \dots, 0)$ and $\nu(y_1)\, \nu(y_2) \dots \nu(y_k) \in S$. Then $\overline{\lim_n}\, A_n \subset M_J^U\, (S) \subset \overline{\lim_n}\, A_n \cup$ (a subset of I); we recall that I is the set of vectors containing only a finite number of positive coordinates

So
$$P\{M_J^U\, (S)\} = P\{\overline{\lim_n}\, A_n\}.$$

Since P is countably additive, is sufficient if we prove that for an arbitrary k, $P(A_k \cup A_{k+1} \cup \dots) \geqslant \underline{\lambda}(S)$. $A_k \cup A_{k+1} \cup \dots$ is the J-magnification of a certain right-complete (with respect to J) set T_k of positive integers. $T_k \cup$ (the set $I_{j_{(k-1)}}$ of positive integers all of whose prime factors are from among $2, 3 \dots, q_{j_{(k-1)}}) \supset S$. Hence
$P(A_k \cup A_{k+1} \cup \dots) = P\{M_J(T_k)\} = \underline{\lambda}(T_k)$ by Lemma 1 $= \underline{\lambda}\{T_k \cup I_{j_{(k-1)}}\} \geqslant \underline{\lambda}\, (S)$.

Theorem 1 : *Let $f(n)$ be a finite real-valued function defined on the set of positive integers. Let $K = \{k_1, k_2, \dots\}$ be an increasing sequence of positive integers. Let $\{C_n\}$, $n \epsilon K$, be a sequence of real numbers such that $f(2^{x_1} \dots q_n^{\ n}) + C_n$ converges with probability 1 to a random variable $g(x)$, as $n \to \infty$ through the numbers in K.*

Then if f has a distribution in the sense of logarithmic density, the sequence C_n converges to a finite limit C and the distribution of f is the same as that of $g(x) - C$.

Proof: Let Q denote the distribution of $f(n)$. The sequence $\{C_n\}$ must be bounded. For, suppose $\{C_n\}$ is unbounded above. Take a small number say 0.01. Let d be a number so large that $P_r\{g(x) > d\} < 0.01$; also let $P_r\{g(x) = d\} = 0$. Let d' be such that $Q(-\infty, d') < 0.02$, $Q(d') = 0$. Since $\{C_n\}$ is unbounded above, we have infinitely many values of $n \,\epsilon\, K$ such that

$$P_r\{f(2^{x_1} \dots q_n^{x_n}) \leqslant d'\} > 0.99.$$

Also, $\underset{n}{E}\{f(n) \leqslant d'\}$ has logarithmic density < 0.02. Now if a set S has logarithmic density < 0.02, then for sufficiently large n, its n-th Π-density ratio $\dfrac{\mu(A_n \cap S)}{\mu(A_n)}$ must be $< 0.02 + (e^\gamma - 1) + \epsilon (\epsilon > 0$ is arbitrarily small$) < 0.92$; here we use Merten's theorem. But $\dfrac{\mu(A_n \cap S)}{\mu(A_n)}$, where $S = \underset{n}{E}\{f(n) \leqslant d'\} = P_r\{f(2^{x_1} \dots q_n^{x_n}) \leqslant d'\} > 0.99$. So we have a contradiction. Similarly $\{C_n\}$ is bounded below.

Now let b be a limit point of $\{C_n\}$; let J be a subsequence of K such that $C_n \to b$ when $n \to \infty$ through J. Then $f(2^{x_1} \ldots q_n^{x_n}) \to g(x) - b$ as $n \to \infty$ through J. Let α be a number such that $P_r\{g(x) - b = \alpha\} = 0$ and $Q(\alpha) = 0$. Then

$$\underset{x}{E}\{g(x)-b<\alpha\} \subset M_J^L\{\underset{n}{E}: f(n)<\alpha\} \text{ and } \underset{x}{E}\{g(x)-b>\alpha\} \subset M_J^L\{\underset{n}{E}: f(n)>\alpha\}.$$

By Lemma 2

$$P_r\{g(x)-b<\alpha\} \leqslant \bar{\lambda}\underset{n}{E}\{f(n)<\alpha\}$$

and

$$P_r\{g(x)-b>\alpha\} \leqslant \bar{\lambda}\underset{n}{E}\{f(n)>\alpha\}.$$

Now we use the hypothesis that f has a distribution in the sense of logarithmic density and deduce that Q is the distribution of $g(x) - b$.

It follows at once that the sequence $\{C_n\}$ has only one limit point.

3. Additive arithmetical functions

A finite real-valued function f defined on the set of positive integers is said to be *additive* in case of $f(mn) = f(m) + f(n)$ whenever m and n are mutually prime. Erdös (1938) proved that the following conditions (we shall refer to them as the EW-conditions) are sufficient to ensure that f has a distribution :

$$\sum_n \frac{f^*(q_n)}{q_n} \text{ and } \sum_n \frac{\{f^*(q_n)\}^2}{q_n} \text{ converge,}$$

here

$$f^*(q_n) = \begin{cases} f(q_n) & \text{if } |f(q_n)| \leqslant 1 \\ 1 & \text{if } |f(q_n)| > 1. \end{cases}$$

Subsequently, Erdös and Wintner (1939) proved that the EW-conditions are necessary if f is to have a distribution. Kac (1949) has pointed out the close analogy between the EW-conditions and Kolmogorov's three series theorem (Doob, 1953, p. 111). Magnification theory enables us to push this analogy into logical deducibility. First we prove the Lemma :

Lemma 3 : *The infinite series* $\Sigma f(q_n^{x_n})$ *converges with probability* 1 *if and only if the EW-conditions hold.*

Proof : Let B be the "box" $(0, 1)X(0, 1)X \ldots$, $B \subset X$. We introduce an auxiliary additive arithmetical function $g(n)$ by putting

$$g(q_n^\alpha) = f(q_n), \quad \alpha = 1, 2, 3, \ldots .$$

By Kolmogorov's three series theorem, validity of the EW-conditions implies the convergence with probability 1 of the series $\sum_n g(q_n^{x_n})$. It follows that the series $\sum_n f(q_n^{x_n})$ converges almost everywhere on B. Now $P(B) = \prod_n (1 - \frac{1}{q_n^2}) > 0$. So by the zero-one law (Doob, 1953, p. 102), $\sum_n f(q_n^{x_n})$ converges with probability 1. The converse is proved similarly, by employing the converse part of the three series theorem.

If the EW-conditions hold, the sequence $f(2^{x_1} \dots q^{x_n})$ converges almost everywhere and hence by Theorem 2 of Paul (1962a, p. 109) f has a distribution in the sense of logarithmic density. This is weaker than the theorem of Erdös who proved that if the EW-conditions hold, f has a distribution in the sense of natural density.

Theorem 2 : *Let $f(n)$ be an additive arithmetical function having a distribution in the sense of logarithmic density. Then the EW-conditions hold.*

Proof : We use the following theorem (see Doob, 1953, p. 121).

Let $y_1, y_2, \dots,$ be mutually independent random variables and suppose that for some $K > 0$

$$\limsup_{n\to\infty} \; P\{|\sum_1^n y_j| \leqslant K\} > 0.$$

Then there is a sequence $\{d_n\}$ of numbers such that $\sum_1^\infty (y_n+d_n)$ converges with probability 1.

Our infinite series $f(2^{x_1}) + f(3^{x_2}) + \dots$ satisfies these conditions. In fact, let Q be the distribution of the function $f(n)$. Let α, β be continuity points of Q such that $Q(\alpha, \beta) = p > 0$. Since $A = E\{n : f(n) \epsilon (\alpha, \beta)\}$ has logarithmic density p, $\underline{\Pi}(A) \geqslant \frac{p}{e^\gamma}$ (γ being Euler's constant). Hence for all sufficiently large n,

$$P_r\{f(2^{x_1} \dots q_n^{x_n}) \epsilon (\alpha, \beta)\} \geqslant \frac{p}{2} > 0.$$

It follows at once from Theorem 1 that the series $f(2^{x_1})+f(3^{x_2})+ \dots$ converges with probability 1; hence the EW-conditions hold.

This theorem was proved by Erdös and Wintner; their theorem is slightly weaker in that they require that $f(n)$ have a distribution in the sense of natural density.

As a corollary to our theorem we have the result :

If an additive arithmetical function has a distribution in the sense of logarithmic density, it has a distribution in the sense of natural density.

Proof : By Theorem 2, the EW-conditions hold; hence by the theorem of Erdös, we get the result.

We can easily generalize the main result above to a class of weakly additive arithmetical functions. Let J consisting of $j_1 < j_2 < \dots$ be arbitrary. We shall call a positive integer greater than 1 a pseudo-prime in case all its prime factors are from among the same block $q_{(j_r+1)}, q_{(j_r+2)}, \dots, q_{j_{(r+1)}}$ (for some r). The pseudo-primes of the form $q_{(j_r+1)}^a \cdots q_{j_{(r+1)}}^d$ will be said to constitute the $(r+1)$-th class of pseudo-primes. Every positive integer has a unique representation as a product of pseudo-primes from different classes (one pseudo-prime at most from each class). We shall refer to this representation as the canonical representation of the number. Two positive

integers m and n will be said to be mutually pseudo-prime in case the class contributing pseudo-prime factors in the canonical representation of m and the corresponding class associated with n are disjoint.

Let $f(n)$ be an arithmetical function that is weakly additive in the sense that $f(mn) = f(m)+f(n)$ whenever m and n are mutually pseudo-prime. Exactly as before, we can prove that if f has a distribution in the sense of logarithmic density, the series

$$\sum_r f\left[q_{(j_r+1)}^{x_{(j_r+1)}} \cdots q_{j_{(r+1)}}^{x_{j_{(r+1)}}}\right]$$

will converge with probability 1.

If $\frac{\log j_{n+1}}{\log j_n}$ is bounded, this condition is also sufficient to ensure that f has a distribution in the sense of logarithmic density. Here we use the generalized magnification theorem (Paul, 1962b).

We recall that a finite collection of arithmetical functions $f_i(m)$, $i = 1, \ldots, r$, will be said to have a joint distribution in case there is a probability distribution Q in r-dimensional space such that for every open set A in r-space with $Q\{Bd(A)\} = 0$, $E_n\{(f_1(n), \ldots, f_r(n)) \in A\}$ has density $Q(A)$; here BdA stands for the boundary of A.

Theorem 3 : *If $f_1(n), \ldots, f_r(n)$ are additive arithmetical functions and each has a distribution (in the sense of logarithmic and hence of natural density), then they have a joint distribution in the sense of logarithmic density.*

Proof : Each of the series $\sum_n f_i(q_n^{x_n})$ converges with probability 1. From this we can deduce our theorem using the magnification inequality $P\{M_L(S)\} \leqslant \underline{\lambda}(S)$, exactly as we proved Theorem 2 of Paul (1962a, p. 109). Similarly, we have

Theorem 4 : *Let J be such that $\frac{\log j_{n+1}}{\log j_n}$ is bounded. Let $f_1(n), \ldots, f_r(n)$ be weakly additive arithmetical functions and let each of these r functions have a distribution in the sense of logarithmic density. Then they have a joint distribution in the sense of logarithmic density.*

REFERENCES

DOOB, J. L. (1953) : *Stochastic Processes*, John Wiley and Sons, New York.

ERDÖS, P. (1938) : On the density of some sequences of numbers. *J. Lond. Math. Soc.*, **13**, 119–127.

——— and WINTNER, A. (1939) : Additive arithmetical functions and statistical independence. *Amer. J. Math.*, **61**, 713–721.

KAC, M. (1949) : Probability methods in some problems of analysis and number theory. *Bull. Amer. Math. Soc.*, **55**, 7, 641–665.

PAUL, E. M. (1962a) : Density in the light of probability theory-I. *Sankhyā*, Series A, **24**, 2, 103–114.

——— (1962b) : Density in the light of probability theory-II. *Sankhyā*, Series A, **24**, 3, 209–212.

TSUJI, M. (1959) : *Potential Theory in Modern Function Theory*, Maruzen and Co., Tokyo.

Paper received : January, 1963.

SOME FIRST PASSAGE PROBLEMS AND THEIR APPLICATION TO QUEUES*

By N. U. PRABHU and U. NARAYAN BHAT

The University of Western Australia

SUMMARY. In this paper we study some first passage problems concerning the process $i+S_n-n$ $(n \geqslant 0)$, where S_n is the sequence of partial sums of independent and identical random variables. The results are applied to Markov chains imbedded in the queueing process $GI|M|1$ and $M|G|1$.

1. Introduction

Let $\{X_n\}$ $(n = 0, 1, 2, \ldots)$, be a sequence of mutually independent and identical random variables which assume non-negative integral values, and $S_n = X_0+X_1+\cdots+X_{n-1}$ $(n = 1, 2, \ldots)$ be the partial sums of $\{X_n\}$. We define the random variables

$$T_i = \min\{n \mid S_n-n \leqslant -i\}, \quad U_j = \min\{n \mid S_n-n \geqslant j\} \qquad \ldots \quad (1.1)$$

$$N = \min\{n \mid S_n-n = 0\} \qquad \ldots \quad (1.2)$$

where $i > 0, j \geqslant 0$; clearly, T_i is the first passage time to a distance i to the left, and U_j the first passage time to a distance j to the right while N is the recurrence time of 0. In this paper we establish some results concerning these random variables. These results are then applied to the queueing systems $GI|M|1$ and $M|G|1$, both of which have a single server and the queue discipline "first come, first served". In the first system the inter-arrival times have the distribution $dF(t)$ $(0 < t < \infty)$, and the service time has the negative exponential distribution $\mu e^{-\mu t}\,dt (0 < t < \infty)$; let Q_n denote the queue length just before the arrival of the n-th customer in this system. In the second system, the arrivals are at random, i.e. the inter-arrival times have the distribution $\mu e^{-\mu t}\,dt$, while the service time has the distribution $dF(t)$; here let Q_n denote the queue length just before the departure of the n-th customer. For a detailed description of the two systems, and references see Kendall (1951, 1953). The process $\{Q_n\}$ $(n = 0, 1, 2, \ldots)$ is in both cases a Markov chain; earlier discussion of these chains was confined mostly to their ergodic behaviour, but during recent years Takács (1960, 1961), Finch (1960), and others have investigated their transient behaviour. However, the methods used so far are sometimes difficult, and moreover, in many cases, the results have not been obtained explicitly. In this paper we deduce all the important properties of $\{Q_n\}$ from those of the random variables T_i, U_j and N.

---

* A summary of the results of this paper was presented at the Second Summer Research Institute of the Australian Mathematical Society held at Canberra in January 1962.

2. Distributions of T_i and U_j

Let

$$\Pr\{X_n = j\} = \begin{cases} k_j & (j = 0, 1, 2, \ldots) \\ 0 & \text{otherwise}; \end{cases} \quad \ldots \quad (2.1)$$

$$K(z) = \sum_{0}^{\infty} k_j z^j (|z| < 1); \; \alpha = E(X_n) < \infty. \quad \ldots \quad (2.2)$$

Further, let

$$k_j^{(n)} = \Pr\{S_n = j\} \;\; (n \geqslant 1), \;\; k_j^{(0)} = 0 \;\; (j \neq 0), \;\; k_0^{(0)} = 1 \; ; \quad \ldots \quad (2.3)$$

$$K_j^{(n)} = \Pr\{S_n \leqslant j\}, \; \alpha_j^{(n)} = \Pr\{S_n > j\}, \; \alpha_j^{(1)} = \alpha_j. \quad \ldots \quad (2.4)$$

We have then the following theorem.

Theorem 1 : *For* $i \geqslant 1$ *we have*

(a) $g(i, n) = \Pr\{T_i = n\} = \Pr\{i+S_r-r > 0 \; (r = 1, 2, \ldots, n-1), \; i+S_n-n = 0\}$

$$= \begin{cases} 0 & \text{if } \; n < i \\ \dfrac{i}{n} k_{n-i}^{(n)} & \text{if } \; n \geqslant i; \end{cases} \quad \ldots \quad (2.5)$$

(b) $$\Pr\{T_i < \infty\} = \sum_{n} g(i, n) = \begin{cases} 1 & \text{if } \alpha < 1 \\ \zeta^i & \text{if } \alpha \geqslant 1 \end{cases} \quad \ldots \quad (2.6)$$

where ζ *is the positive root of the equation* $z = K(z)$; *and*

(c) $E(T_i) = i(1-\alpha)^{-1}$ *if* $\alpha < 1, = \infty$ *if* $\alpha = 1$. $\quad \ldots \quad (2.7)$

If we interpret X_n as the amount of water which flows into a dam during a unit time interval, and suppose that, as in Moran's (1959) storage model, a unit amount of water is released from the dam at the end of each such interval, unless the dam is empty, then T_i is the duration of the "wet period", i.e. the time taken by the dam with initial content i to dry up. The above results are a restatement of those derived by Kendall (1957) for the analogous problem in continuous time; however, for the sake of completeness the main outlines of the proof are given below. Since $i+S_r-r > 0$ for $r < i$ we must have $T_i \geqslant i$; moreover, since we can write $i+S_n-n = S_i+S_{n-i}-(n-i) \; (n \geqslant i)$, it follows that

$$T_i = i+T(S_i). \quad \ldots \quad (2.8)$$

Using (2.8) we obtain a difference equation for $g(i, n)$, which yields the solution (2.5) [see also Gani (1958) and Takács (1961)]. Further, it is obvious that $\Pr\{T_i < \infty\}$ is of the form ζ^i, so that (2.8) gives

$$\zeta^i = \sum_{0}^{\infty} k_j^{(i)} \Pr\{T_j < \infty\} = [K(\zeta)]^i \quad \ldots \quad (2.9)$$

which leads to (b). Finally, using $E(T_i) = iE(T_1)$, we obtain from (2.8),

$$E(T_1 = 1+\sum_{0}^{\infty} k_j E(T_j) = 1+\alpha E(T_1) \quad \ldots \quad (2.10)$$

which gives (c).

Theorem 2 : *For $i \geqslant 1$ we have*

$$\Pr\{i+S_r-r>0 \quad (r=1,2,\ldots,n-1);\quad i+S_n-n=j\}$$

$$=k^{(n)}_{n+j-i}-\sum_{m=1}^{n-1} g(i,n-m)\,k^{(m)}_{m+j}. \qquad \ldots \quad (2.11)$$

For, we have

$$k^{(n)}_{n+j-i}=\Pr\{i+S_n-n=j\}=\Pr\{T_i \geqslant n,\, i+S_n-n=j\}$$

$$+\Pr\{T_i<n;\quad i+S_n-n=j\}. \qquad \ldots \quad (2.12)$$

The first term in the last expression is the required probability, while the second term can be written as

$$\sum_{m=1}^{n-1}\Pr\{T_i=m\}\,\Pr\{i+S_n-n=j\,|\,T_i=m\}$$

$$=\sum_{m=1}^{n-1} g(i,m)\,\Pr\{i+S_n-n=j\,|\,i+S_m-m=0\}$$

$$=\sum_{=1}^{n-1} g(i,m)\,k^{(n-m)}_{n-m+j} \qquad \ldots \quad (2.13)$$

which proves (2.11).

We observe that when $j=0$ the expression (2.11) reduces to $g(i,n)$,

since
$$k^{(n)}_{n-i}=g(i,n)+\sum_{m=1}^{n-1} g(i,n-m)\,k^{(m)}_{m}, \qquad \ldots \quad (2.14)$$

while, when $j<0$, (2.11) vanishes identically, as it should,

since
$$k^{(n)}_{n-\nu}=\sum_{m=1}^{n-1} g(i,n-m)\,k^{(m)}_{m-\nu} \quad (\nu>0). \qquad \ldots \quad (2.15)$$

The identities (2.14) and (2.15) can be easily proved. Thus Theorem 2 is a generalization of (2.5).

Theorem 3 : *We have*

(a) $\Pr\{U_0=1\}=\alpha_0$

$$\Pr\{U_0=n+1\}=\sum_{i=1}^{\infty}\alpha_i\, g(i,n) \qquad (n=1,2,\ldots); \qquad \ldots \quad (2.16)$$

(b) $$\Pr\{U_0<\infty\}=\begin{cases}1 & \text{if } \alpha \geqslant 1\\ \alpha & \text{if } \alpha<1.\end{cases} \qquad \ldots \quad (2.17)$$

For, we have

$$\Pr\{U_0 = 1\} = \Pr\{S_1 - 1 \geqslant 0\} = \alpha_0, \qquad \ldots \quad (2.18)$$

and for $n \geqslant 1$,

$$\Pr\{U_0 = n+1\} = \Pr\{S_r - r < 0 \quad (r = 1, 2, \ldots, n), \quad S_{n+1} - (n+1) \geqslant 0\}$$

$$= \sum_{i=1}^{\infty} \Pr\{S_r - r < 0 \ (r = 1, 2, \ldots, n-1), \ S_n - n = -i\}$$
$$\times \Pr\{S_{n+1} - (n+1) \geqslant 0 \mid S_n - n = -i\}$$

$$= \sum_{i=1}^{\infty} \Pr\{S_{n-r} - (n-r) = S_n - n - (S_r - r) > -i \, (r = 1, 2, \ldots, n-1), \ S_n - n = -i\}$$
$$\times \Pr\{S_{n+1} - (n+1) - (S_n - n + i) \geqslant 0\}$$

$$= \sum_{i=1}^{\infty} \alpha_i \Pr\{i + S_r - r > 0 \ (r = 1, 2, \ldots, n-1), \ i + S_n - n = 0\} \qquad \ldots \quad (2.19)$$

and the result now follows from (2.5). Further, we have

$$\Pr\{U_0 < \infty\} = \sum_{0}^{\infty} \Pr\{U_0 = n\} = \alpha_0 + \sum_{i=1}^{\infty} \alpha_i \Pr\{T_i < \infty\}$$

$$= \begin{cases} \sum_{0}^{\infty} \alpha_i = \alpha & \text{if } \alpha < 1 \\ \sum_{0}^{\infty} \alpha_i \zeta^i = \dfrac{1 - K(\zeta)}{1 - \zeta} = 1 & \text{if } \alpha \geqslant 1, \end{cases} \qquad \ldots \quad (2.20)$$

using (2.6).

Theorem 4 : *For $j \geqslant 1$ we have*

(a) $\Pr\{U_j = 1\} = \alpha_j$

$$\Pr\{U_j = n+1\} = \sum_{i=1}^{\infty} \alpha_i \{k^{(n)}_{n+j-i} - \sum_{m=1}^{n-1} g(i, n-m) k^{(m)}_{m+j}\} \ (n \geqslant 1) \qquad \ldots \quad (2.21)$$

(b) $$\Pr\{U_j < \infty\} = \begin{cases} 1 & \text{if } \alpha \geqslant 1 \\ (1-\alpha) \sum_{n=0}^{\infty} k^{(n)}_{n+j} & \text{if } \alpha < 1. \end{cases} \qquad \ldots \quad (2.22)$$

The proof of (a) is similar to that of (2.16), but we have now to apply (2.11). To prove (b) we note that

$$\Pr\{U_j < \infty\} = \alpha_j + \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} \alpha_i k^{(n)}_{n+j-i} - \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} \sum_{m=1}^{n-1} \alpha_i k^{(m)}_{m+j} g(i, n-m)$$

$$= \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} \alpha_i k^{(n)}_{n+j-i} - \sum_{n=0}^{\infty} k^{(n)}_{n+j} \quad \Pr\{U_0 < \infty\}. \qquad \ldots \quad (2.23)$$

This last expression is the coefficient of z^j in the formal expansion of

$$\left[\sum_{0}^{\infty} \alpha_i z^i - 1 + \Pr\{U_0 = \infty\}\right] \sum_{0}^{\infty} \left[\frac{K(z)}{z}\right]^n. \qquad \ldots \quad (2.24)$$

By a familiar type of argument [see for instance, Prabhu (1960)], it can be proved that $|K(z)| < |z|$ in the region $1 < |z| < \zeta$ if $\alpha < 1$ and in $\zeta < |z| < 1$ if $\alpha \geqslant 1$; moreover $\Sigma\, \alpha_i z^i = [1-K(z)]\,(1-z)^{-1}$ in both cases, if we assume that $K(z)$ is capable of an analytic extension to the region $1 < |z| < \zeta$ in the case $\alpha < 1$. Thus the expression (2.24) reduces to

$$\frac{z}{z-K(z)}\left[\frac{z-K(z)}{1-z}+\Pr\{U_0=\infty\}\right]=\begin{cases}\dfrac{z}{1-z} & (\zeta<|z|<1, \quad \alpha\geqslant 1)\\[2ex] \dfrac{z}{1-z}+(1-\alpha)\dfrac{z}{z-K(z)} & (1<|z|<\zeta, \alpha<1).\end{cases} \quad \ldots (2.25)$$

Expanding the functions in the regions of their validity we find from (2.25) that the coefficient of z^j is as given in (2.22).

We also have

$$\Pr\{U_j = n+1\} = \sum_{i=1}^{\infty} \Pr\{S_1-1 = j-i,\ S_r-r<j\ (r=2,3,\ldots,n),\ S_{n+1}-(n+1)\geqslant j\}$$

$$= \sum_{=1}^{\infty} k_{j+1-i} \Pr\{S_r-r<j\ (r=2,3,\ldots,n),$$
$$S_{n+1}-(n+1)\geqslant j \mid S_1-1=j-i\}$$

$$= \sum_{i=1}^{\infty} k_{j+1-i}\ \Pr\{U_i=n\} \quad (n\geqslant 1, j\geqslant 0). \quad \ldots (2.26)$$

Adding up (2.26) over $n = 1, 2, \ldots$ we obtain

$$\Pr\{U_j<\infty\} = \sum_{i=1}^{\infty} k_{j+1-i}\ \Pr\{U_i<\infty\}+\alpha_j \quad (j\geqslant 0)$$

whence we get the result

$$\sum_{0}^{\infty} z^j\ \Pr\{U_j=\infty\} = \frac{K(z)}{K(z)-z}\Pr\{U_0=\infty\},\ |z|<1. \quad \ldots (2.27)$$

3. The Markov chain $Z_n = i+S_n-n$

It is clear that the process $Z_n = i+S_n-n$, $Z_0 = i$, is a Markov chain, but we have not made use of any special property of this chain in deriving the results of Section 2. The transition probabilities of $\{Z_n\}$ are given by

$$P_{ij}^{(n)} = \Pr\{Z_n=j\mid Z_0=i\} = k_{n+j-i}^{(n)}\ (n\geqslant 1;\ i,j=\cdots-1,0,1,2,\ldots); \quad \ldots (3.1)$$

$$P_{ij}^{(0)} = k_{j-i}^{(0)} = 0 \quad (j\neq i), = 1 \quad (j=i). \quad \ldots (3.2)$$

The random variable N defined by (1.2) is the recurrence time of the state 0 of the chain $\{Z_n\}$. The following results are implicit in the more general results derived by Kemperman (1961) using complex variable methods, but in their explicit form are a consequence of the results of Section 2.

Theorem 5 : *Let* $F_{00}^{(n)} = \Pr\{N = n\}$; *we have then*

(a) $F_{00}^{(1)} = k_1$

$$F_{00}^{(n+1)} = \sum_{i=2}^{\infty} ik_i \, g(i-1, n) \quad (n \geqslant 1). \qquad \ldots \quad (3.3)$$

(b) $$F_{00} = \Pr\{N < \infty\} = \begin{cases} \alpha & \text{if } \alpha < 1 \\ K'(\zeta) < 1 & \text{if } \alpha \geqslant 1. \end{cases} \qquad \ldots \quad (3.4)$$

For, $$F_{00}^{(1)} = \Pr\{S_1 - 1 = 0\} = k_1,$$

while for $n \geqslant 1$ we have

$$\begin{aligned} F_{00}^{(n+1)} &= \Pr\{S_r - r \neq 0 \ (r = 1, 2, \ldots, n),\ S_{n+1}-(n+1) = 0\} \\ &= \Pr\{S_r - r > 0 \ \ (r = 1, 2, \ldots, n),\ S_{n+1}-(n+1) = 0\} \\ &\quad + \Pr\{S_r - r < 0 \ (r = 1, 2, \ldots, n),\ \ S_{n+1}-(n+1) = 0\} \\ &\quad + \sum_{m=1}^{n-1} \Pr\{S_r - r < 0 \ (r = 1, 2, \ldots, m), \\ &\qquad\qquad S_r - r > 0 \ (r = m+1, \ldots, n),\ S_{n+1}-(n+1) = 0\}. \end{aligned} \qquad \ldots \quad (3.5)$$

Since $S_r - r = S_{n+1}-(n+1) - S_{n+1-r}-(n+1-r)$ it is clear that the first term on the right hand side of (3.5) is equal to the second; the latter, however, is

$$= \sum_{i=1}^{\infty} \Pr\{S_r - r < 0 \ (r = 1, 2, \ldots, n-1),\ S_n - n = -i,\ S_{n+1}-(n+1) = 0\}$$

$$= \sum_{i=1}^{\infty} \Pr\{i + S_{n-r} - (n-r) > 0 \ (r = 1, 2, \ldots, n-1), \quad i + S_n - n = 0\}.$$

$$\Pr\{S_{n+1}-(n+1) = 0 \mid s_n - n = -i\} = \sum_{i=1}^{\infty} k_{i+1} \, g(i, n). \qquad \ldots \quad (3.6)$$

The remaining term on the right hand side of (3.5) is

$$= \sum_{m=1}^{n-1} \sum_{i=1}^{\infty} \Pr\{S_r - r < 0 \ (r = 1, 2, \ldots, m),\ S_{m+1}-(m+1) = i\}.$$

$$\Pr\{S_r - r > 0 \ (r = m+2, \ldots, n),\ S_{n+1}-(n+1) = 0 \mid S_{m+1}-(m+1) = i\}$$

$$= \sum_{m=1}^{n-1} \sum_{i=1}^{\infty} g(i, n-m) \sum_{j=1}^{\infty} \Pr\{S_r - r < 0 \ (r = 1, 2, \ldots, m), S_m - m = -j,\ S_{m+1}-(m+1) = i\}$$

$$= \sum_{m=1}^{n-1} \sum_{i=1}^{\infty} g(i, n-m) \sum_{j=1}^{\infty} k_{j+i+1} \, g(j, m) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} k_{j+i+1} \, g(j+i, n) \qquad \ldots \quad (3.7)$$

$$= \sum_{\nu=1}^{\infty} (\nu-2) k_\nu \, g(\nu-1, n) \qquad \ldots \quad (3.8)$$

where in (3.7) we have used the obvious property $T_i + T_j = T_{i+j}$. From (3.6) and (3.8) we obtain (3.3) for $n \geqslant 1$. Further,

we have $$F_{00} = \sum_{1}^{\infty} F_{00}^{(n)} = k_1 + \sum_{2}^{\infty} i\, k_i \Pr\{T_{i-1} < \infty\}$$

$$= \begin{cases} \alpha & \text{if } \alpha \leqslant 1 \\ \sum_{1}^{\infty} ik_i \zeta^{i-1} = K'(\zeta) < 1 & \text{if } \alpha > 1. \end{cases} \qquad \ldots \quad (3.9)$$

From (3.4) we see that the Markov chain $\{Z_n\}$ is transient except when $\alpha = 1$; moreover, the mean recurrence time of 0 is $E(N) = 1+\sum_{2}^{\infty} ik_i \ E(T_{i-1}) = \infty$ when $\alpha = 1$, so that in this case the chain is persistent null. It follows from (3.1) that the series $\sum_{n=0}^{\infty} k_{n+j}^{(n)}$ converges except when $\alpha = 1$. More specifically we have the following theorem.

Theorem 6 : *If* $\alpha \neq 1$ *we have*

$$\text{(a)} \sum_{n=i}^{\infty} k_{n-i}^{(n)} = \begin{cases} (1-\alpha)^{-1} & \text{if } \alpha < 1 \\ \zeta^i[1-K'(\zeta)]^{-1} & \text{if } \alpha > 1 \end{cases} \quad (i \geqslant 0), \qquad \dots \quad (3.10)$$

$$\text{(b)} \sum_{n=0}^{\infty} k_{n+j}^{(n)} < \begin{cases} (1-\alpha)^{-1} & \text{if } \alpha < 1 \\ [1-K'(\zeta)]^{-1} & \text{if } \alpha > 1 \end{cases} \quad (j > 0). \qquad \dots \quad (3.11)$$

Incidentally, $\Pr\{U_j < \infty\} < 1$ if $\alpha < 1$.

From the familiar identity $F_{ij} = \sum_{1}^{\infty} P_{ij}^{(n)} / \sum_{0}^{\infty} P_{jj}^{(n)} \quad (\alpha \neq 1)$ ... (3.12)

we obtain the following for special values of i, j :

(i) Let $i = 0, j = 0$; then $\sum_{0}^{\infty} P_{00}^{(n)} = (1-F_{00})^{-1}$ and using (3.4) we obtain (a) with $i = 0$.

(ii) Let $i > 0$, $j = 0$; then $\sum_{1}^{\infty} P_{i0}^{(n)} = F_{i0} \sum_{0}^{\infty} k_n^{(n)}$, and, since $F_{i0} = \Pr\{T_i < \infty\}$ we obtain (a) for $i > 0$.

(iii) Let $i = 0$, $j > 0$; then $\sum_{1}^{\infty} P_{0j}^{(n)} = F_{0j} \sum_{0}^{\infty} k_n^{(n)} < \sum_{0}^{\infty} k_n^{(n)}$ since $F_{0j} < 1$ for $\alpha \neq 1$; this gives (b).

4. The queue $GI|M|1$

Let $t_0, t_1, t_2, \dots$ be the epochs of arrival of the successive customers in the system, and Q_n denote the queue length at $t = t_n - 0$. Also, let X_n be the number of departures during the interval $(t_n, t_{n+1}-0)$ $(n = 0, 1, 2, \dots)$; clearly $X_0, X_1, \dots$ are mutually independent and identical random variables, with

$$k_j = \Pr\{X_n = j\} = \int_0^\infty e^{-\mu t} \frac{(\mu t)^j}{j\,!} \, dF(t) \quad (j \geqslant 0). \qquad \dots \quad (4.1)$$

We have $K(z) = \sum_{0}^{\infty} k_j z^j = \psi(\mu - \mu z), \quad \alpha = E(X_n) = \rho^{-1},$ ... (4.2)

where $\psi(\theta)$ is the Laplace transform (L.T.) of $dF(t)$ and ρ is the relative traffic intensity; further,

$$k_j^{(n)} = \int_0^\infty e^{-\mu t} \frac{(\mu t)^j}{j\,!} \, dF_n(t) \qquad \dots \quad (4.3)$$

$$\alpha_j^{(n)} = \int_0^\infty e^{-\mu t} \mu^{j+1} \frac{t^j}{j!} [1-F_n(t)]dt, \; K_j^{(n)} = \int_0^\infty e^{-\mu t} \mu^{j+1} \frac{t^j}{j!} F_n(t) dt, \qquad \dots \quad (4.4)$$

where $F_n(t)$ is the n-fold convolution of $F(t)$ with itself, and $F_0(t) = 0$ if $t < 0$ and $= 1$ if $t \geqslant 0$.

For the Markov chain $\{Q_n\}$ we have the recurrence relations

$$Q_{n+1} = \begin{cases} Q_n+1-X_n & \text{if } Q_n+1-X_n > 0 \\ 0 & \text{if } Q_n+1-X_n \leqslant 0, \end{cases} \qquad \ldots \quad (4.5)$$

whence we obtain

$$Q_n = \max\{r-S_r \quad (r = 0, 1, 2, \ldots, n-1), \quad i+n-S_n\}, \qquad \ldots \quad (4.6)$$

where we have denoted $Q_0 = i$. The transition probabilities of $\{Q_n\}$ are therefore given by

$$\Pr\{Q_n < j \mid Q_0 = i\} = \Pr\{r-S_r < j \ (r = 0, 1, \ldots, n-1), \ i+n-S_n < j\}$$

$$= \sum_{\nu = +1}^{\infty} \Pr\{j+S_r-r > 0 \ (r = 0, 1, \ldots, n-1), \ j+S_n-n = \nu\}$$

$$= \sum_{i+1}^{\infty} \{k^{(n)}_{n+\nu-j} - \sum_{m=1}^{n-1} g(j, n-m) \ k^{(m)}_{m+\nu}\} \text{ from (2.11)}$$

$$= \alpha^{(n)}_{n-j+i} - \sum_{m=1}^{n-1} g(j, n-m) \, \alpha^{(m)}_{m+i}, \qquad \ldots \quad (4.7)$$

a result which has not been explicitly obtained before. In particular,

$$\Pr\{Q_n < j \mid Q_0 = 0\} = 1 - \sum_{m=1}^{n} g(j, m). \qquad \ldots \quad (4.8)$$

We define the zero-avoiding transition probabilities ${}^0P^{(n)}_{ij}$ of $\{Q_n\}$ as follows :

$${}^0P^{(n)}_{ij} = \Pr\{Q_r > 0 \quad (r = 1, 2, \ldots, n-1), Q_n = j \mid Q_0 = i\}. \qquad \ldots \quad (4.9)$$

For $j > 0$ we have

$${}^0P^{(n)}_{ij} = \Pr\{i+r-S_r > 0 \ (r = 1, 2, \ldots, n-1); i+n-S_n = j\}$$

$$= \Pr\{(n-r)-S_{n-r} < j \ (r = 1, 2, \ldots, n-1), \quad j+S_n-n = i\}$$

$$= \Pr\{j+S_r-r > 0 \ (r = 1, 2, \ldots, n-1), \quad j+S_n-n = i\}$$

$$= \begin{cases} g(j, n) & \text{if } \ i = 0 \\ k^{(n)}_{n+i-j} - \sum\limits_{m=1}^{n-1} g(j, n-m) \ k^{(m)}_{m+i} & \text{if } \ i > 0, \end{cases} \qquad \ldots \quad (4.10)$$

while, for $j = 0$ we have

$${}^0P^{(n)}_{i0} = \Pr\{i+r-S_r > 0 \quad (r = 1, 2, \ldots, n-1), \quad i+n-S_n \leqslant 0\}$$

$$= \Pr\{S_r-r < i \ (r = 1, 2, \ldots, n-1), \quad S_n-n \geqslant i\}$$

$$= \Pr\{U_i = n\} \quad (i \geqslant 0). \qquad \ldots \quad (4.11)$$

In particular, ${}^0P^{(n)}_{00}$ $(n = 1, 2, \ldots)$ is the distribution of the number of customers served during a busy period, and is therefore given by (2.16), which agrees with the result of Takács (1961, p. 406). It follows that the chain $\{Q_n\}$ is persistent null if $\rho = 1$, persistent non-null if $\rho < 1$, and transient if $\rho > 1$. The limiting queue length is given by

$$Q_\infty = \lim_{n\to\infty} Q_n = \max_{r \geqslant 0} (r - S_r), \qquad \ldots \quad (4.12)$$

when we obtain

$$\Pr\{Q_\infty < j) = \Pr\{\max_{r\geqslant 0} (r - S_r) < j\} = \Pr\{j + S_r - r > 0,\ r \geqslant 0\}$$

$$= \Pr\{T_j = \infty\} = \begin{cases} 0 & \text{if } \rho \geqslant 1 \\ 1 - \zeta^j & \text{if } \rho < 1, \end{cases} \qquad \ldots \quad (4.13)$$

which gives the familiar result for the stationary distribution of the queue-length.

5. The queue $M|G|1$

Here, let $t_0, t_1, t_2, \ldots$ denote the epochs of successive departures, and Q_n the queue length at time $t = t_n + 0$. Further, let X_n be the number of arrivals during $(t_n + 0,\ t_{n+1})$ $(n = 0, 1, \ldots)$; then $X_0, X_1, X_2, \ldots$ are mutually independent and identical random variables, with the distribution (4.1), and $\alpha = E(X_n) = \rho$, the relative traffic intensity.

We have
$$Q_{n+1} = \begin{cases} Q_n - 1 + X_n & \text{if } Q_n > 0 \\ X_n & \text{if } Q_n = 0, \end{cases} \qquad \ldots \quad (5.1)$$

whence we obtain

$$Q_n = \max\{X_{n-1} + X_{n-2} + \cdots + X_{n-r} - r + 1 \ (1 \leqslant r \leqslant n-1),\ i + S_n - n\},$$

where $Q_0 = i$.

Since
$$X_{n-1} + X_{n-2} + \cdots + X_{n-r} = S_n - S_{n-r} = S_r$$

we can write

$$Q_n = \max\{S_r - r + 1 \ (r = 1, 2, \ldots, n-1);\ i + S_n - n\}. \qquad \ldots \quad (5.2)$$

For the limiting queue length we have

$$Q_\infty = \max_{r \geqslant 1} (S_r - r + 1)\ ; \qquad \ldots \quad (5.3)$$

the limiting distribution of queue length is therefore given by

$$\Pr\{Q_\infty \leqslant j\} = \Pr\{\max_{r\geqslant 1} (S_r - r + 1) \leqslant j\} = \Pr\{S_r - r < j,\ r \geqslant 1\}$$

$$= \Pr\{U_j = \infty\}. \qquad \ldots \quad (5.4)$$

Hence we obtain

$$\Pr\{Q_\infty = 0\} = \begin{cases} 0 & \text{if } \rho \geqslant 1 \\ 1 - \rho & \text{if } \rho < 1 \end{cases} \qquad \ldots \quad (5.5)$$

and
$$\Pr\{Q_\infty \leqslant j\} = \begin{cases} 0 & \text{if } \rho \geqslant 1 \\ 1 - (1-\rho)\sum_0^\infty k^{(n)}_{n+j} & \text{if } \rho < 1, \end{cases} \qquad \ldots \quad (5.6)$$

which is essentially the result proved by Finch (1960) by incomplete arguments. We also note that when $\rho < 1$, we have the well-known result

$$\sum_0^{\infty} z^j \Pr\{Q_\infty \leqslant j\} = (1-\rho)\frac{K(z)}{K(z)-z} \quad (|z| < 1) \qquad \ldots \quad (5.7)$$

as a consequence of (2.27); expanding (5.7) as a power series in z we obtain

$$\Pr\{Q_\infty \leqslant j\} = (1-\rho)\sum_{n=0}^{j} k_{j-n}^{(-n)} \qquad \ldots \quad (5.8)$$

where $k_j^{(-n)}$ is the coefficient of z^j in the expansion of $K(z)^{-n}$; this alternative expression may be more useful in particular cases.

The transition probabilities of $\{Q_n\}$ are given by

$$\begin{aligned}\Pr\{Q_n \leqslant j \mid Q_0 = i\} &= \Pr\{S_r - r + 1 \leqslant j \ (r = 1, 2, \ldots, n-1),\ i + S_n - n \leqslant j\} \\ &= \sum_{\nu=i}^{\infty} \Pr\{S_r - r < j \quad (r = 1, 2, \ldots, n-1),\ S_n - n = j - \nu\} \\ &= \sum_{\nu=i}^{\infty} \Pr\{S_{n-r} - (n-r) > -\nu (r = 1, 2, \ldots, n-1),\ S_n - n = j - \nu\} \\ &= \sum_{\nu=i}^{\infty} \Pr\{\nu + S_r - r > 0 \ (r = 1, 2, \ldots, n-1),\ \nu + S_n - n = j\}. \qquad \ldots \quad (5.9)\end{aligned}$$

Hence, in particular when $j = 0$ we obtain

$$P_{i0}^{(n)} = \Pr\{Q_n = 0 \mid Q_0 = i\} = \sum_{\nu=i}^{\infty} g(\nu, n), \qquad \ldots \quad (5.10)$$

and for $j \geqslant 0$,

$$\begin{aligned}\Pr\{Q_n \leqslant j \mid Q_0 = i\} &= \sum_{\nu=i}^{\infty} \{k_{n+j-\nu}^{(n)} - \sum_{m=1}^{n-1} g(\nu, n-m)\, k_{m+j}^{(m)}\} \\ &= K_{n+j-i}^{(n)} - \sum_{m=1}^{n-1} P_{i0}^{(n-m)} k_{m+j}^{(m)}, \qquad \ldots \quad (5.11)\end{aligned}$$

[cf. Finch (1960), equation 16, and also in another context Yeo (1961). The present methods seem to be much more straightforward].

The zero-avoiding transition probabilities defined as in (4.9) are given by

$${}^0P_{ij}^{(n)} = \begin{cases} \Pr\{i + S_r - r > 0 \ (r = 1, 2, \ldots, n-1),\ i + S_n - n = j\} \ (i \geqslant 1) \\ \Pr\{1 + S_r - r > 0 \ (r = 1, 2, \ldots, n-1),\ 1 + S_n - n = j\} \ (i = 0) \end{cases} \qquad \ldots \quad (5.12)$$

and these are given by (2.11). In particular,

$${}^0P_{i0}^{(n)} = g(i, n) = \frac{i}{n}\int_0^{\infty} e^{-\mu t}\frac{(\mu t)^{n-i}}{(n-i)!}\, dF_n(t) \quad (n \geqslant i) \qquad \ldots \quad (5.13)$$

gives the distribution of the number of customers served during a busy period initiated by i customers.

6. Particular cases

The queue $D|M|1$. Let $F(t) = 0$ if $t < \lambda$, and $= 1$ if $t \geqslant \lambda$. Then $K(z)$ $e^{-(1-z)/\rho}$, so that

$$\alpha_j^{(n)} = \sum_{r=j+1}^{\infty} e^{-n/\rho} \frac{(n/\rho)^r}{r\,!} r!,\; g(j, n) = \frac{j}{n} e^{-n/\rho} \frac{(n/\rho)^{n-j}}{(n-j)\,!} \qquad \dots \quad (6.1)$$

Hence

$$\Pr\{Q_n < j \mid Q_0 = i\} = \sum_{s=n-j+i+1} e^{-n/\rho} \frac{\rho^{-s}}{s!} \{n^s - \beta_j^{(n)}(s)\}, \qquad \dots \quad (6.2)$$

where

$$\beta_j^{(n)}(s) = \begin{cases} 0 & \text{if } n < j \\ \displaystyle\sum_{m=0}^{n-} \frac{j}{m+j}\binom{s}{m}(m+j)^m(n-j-m)^{s-m} & \text{if } n \geqslant j. \end{cases} \qquad \dots \quad (6.3)$$

The queue $E_k|M|1$. Let $dF(t) = (k\lambda^{-1})^k e^{-kt/\lambda} t^{k-1} dt/(k-1)!$ $(0 < t < \infty)$; then $K(z) = a^k(1-bz)^{-k}$ where $b = (1+k\rho)^{-1}$ and $a = 1-b$.

Hence
$$\alpha_j^{(n)} = \sum_{r=j+1}^{\infty} \binom{-nk}{r} a^{nk}(-b)^r,\; g(j, n) = \frac{j}{n}\binom{-nk}{n-j} a^{nk}(-b)^{n-j}, \qquad \dots \quad (6.4)$$

and

$$\Pr\{Q_n < j | Q_0 = i\}$$
$$= a^{nk} \sum_{s=n-j+i+1}^{\infty} (-b)^s \left\{ \binom{-nk}{s} - \sum_{m=0}^{s} \frac{j}{m+j} \binom{-jk-mk}{m} \binom{-nk+jk+mk}{s-m} \right\} \quad (6.5)$$

The queue $M|D|1$. Let the service time distribution be given by $F(t) = 0$ if $t < \lambda$, and $=1$ if $t \geqslant \lambda$. Then $K(z) = e^{-\rho(1-z)}$, and hence

$$P_{i0}^{(n)} = \sum_{i}^{\infty} \frac{\nu e^{-n\rho}(n\rho)^{u-\nu}}{n(n-\nu)\,!} \qquad \dots \quad (6.6)$$

and

$$\Pr\{Q_n \leqslant j | Q_0 = i\}$$
$$= \sum_{0}^{n+j-i} e^{-n\rho} \frac{(n\rho)^s}{s!} - \sum_{m=i}^{n} e^{-(n-m)\rho} \frac{(n\rho - m\rho)^{n-m+j}}{(n-m+j)!} \sum_{\nu=i}^{m} \frac{\nu e^{-m\rho}}{m(n-\nu)!} (m\rho)^{m-\nu}$$
$$= \sum_{s=0}^{n+j-i} e^{-n\rho} \frac{\rho^s}{s!} [n^s - \beta_{n+j-s}^{(n)}(s)]. \qquad \dots \quad (6.7)$$

The limiting distribution of the queue length is given by

$$\Pr\{Q_\infty \leqslant j\} = (1-\rho) \sum_{n=0}^{j} e^{n\rho} \frac{(-n\rho)^{j-n}}{(j-n)\,!} \quad (\rho < 1) \qquad \dots \quad (6.8)$$

[cf. Syski (1960), p. 325].

The queue $M|E_k|1$. Let the service time distribution be given by $dF(t) = (k\,\lambda^{-1})^k e^{-kt/\lambda} t^{k-1} dt/(k-1)!$ $(0 < t < \infty)$; then $K(z) = a^k (1-bz)^{-k}$, where $a = k(k+\rho)^{-1}$ and $b = 1-a$.

Hence
$$P_{i0}^{(n)} = \sum_{\nu=i}^{n} \frac{\nu}{n} \binom{-nk}{n-\nu} a^{nk} (-b)^{n-\nu}, \qquad \ldots \quad (6.9)$$

and
$$\Pr\{Q_n \leqslant j \mid Q_0 = i\} = a^{nk} \sum_{s=0}^{n+j-i} \binom{-nk}{s} (-b)^s$$
$$-a^{nk} \sum_{s=j}^{n+j-i} (-b)^s \sum_{m=0}^{s-j} \frac{n+j-s}{n+j-s+m} \binom{-nk-jk+sk-mk}{m} \binom{mk+jk-sk}{s-m} \quad \ldots \quad (6.10)$$

The limiting distribution of queue length is given by
$$\Pr\{Q_\infty \leqslant j\} = (1-\rho) \sum_{n=0}^{j} a^{-nk} \binom{nk}{j-n} (-b)^{j-n} \qquad (\rho < 1). \qquad \ldots \quad (6.11)$$

The usual expression for the limiting distribution in this case is a weighted sum of k geometric terms, the common ratios and the weights being obtained from a certain characteristic equation [see Syski (1960), p. 321]; however, the result (6.11) appears to be much simpler.

Putting $k = 1$ in (6.10) we obtain the transition probabilities of $\{Q_n\}$ for the queue $M|M|1$ as
$$\Pr\{Q_n \leqslant j \mid Q_0 = i\} = a^n \sum_{s=0}^{n+j-i} \binom{-n}{s} (-b)^s$$
$$-a^n \sum_{s=j}^{n+j-i} (-b)^s \sum_{m=0}^{s-j} \frac{n+j-s}{n+j-s+m} \binom{-n-j+s-m}{m} \binom{m+j-s}{s-m} \quad \ldots \quad (6.12)$$
where $a = (1+\rho)^{-1}$ and $b = 1-a$ (cf. Finch, 1960, equation (33)).

References

FINCH, P. D. (1960) : On the transient behaviour of a simple queue. *J. Roy. Stat. Soc.*, B, **22**, 277–284.

GANI, J. (1958) : Elementary methods in an occupancy problem of storage. *Math. Ann.*, **136**, 454–465.

KEMPERMAN, J. H. B. (1961) : *The Passage Problem for a Stationary Markov Chain.* University of Chicago Press.

KENDALL, D. G. (1951) : Some problems in the theory of queues. *J. Roy. Stat. Soc.*, B, **13**, 151–185.

——— (1953) : Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *Ann. Math. Stat.*, **24**, 338–354.

——— (1957) : Some problems in the theory of dams. *J. Roy. Stat. Soc.*, B, **19**, 207–212.

MORAN, P. A. P. (1959) : *The Theory of Storage*, Methuen, London.

PRABHU, N. U. (1960) : Application of storage theory to queues with Poisson arrivals. *Ann. Math. Stat.*, **31**, 475–482.

SYSKI, R. (1960) : *Introduction to Congestion Theory in Telephone Systems*, Oliver and Boyd, London.

TAKÁCS, L. (1960) : The transient behaviour of a single server queueing process with recurrent input and exponentially distributed service time. *J. Oper. Res.*, **8**, 231–245.

——— (1961) : The probability law of the busy period for two types of queueing processes. *J. Oper. Res.*, **9**, 402–407.

YEO, G. (1961) : The time dependent solution for an infinite dam with discrete additive inputs. *J. Roy. Stat. Soc.*, B, **23**, 173–179.

Paper received : June, 1962.

ON STABLE SEQUENCES OF EVENTS

By ALFRED RÉNYI

Mathematical Institute, Hungarian Academy of Sciences

SUMMARY. A sequence $\{A_n\}$ of events is called a stable sequence if for every event B the limit $\lim_{n \to +\infty} P(A_nB) = Q(B)$ exists. It is shown that in this case Q is a bounded measure which is absolutely continuous with respect to the underlying probability measure P. The Radon-Nikodym derivative $\frac{dQ}{dP} = \alpha$ is called the local density of the stable sequence $\{A_n\}$. Criteria for a sequence of events being stable are given, further examples of stable sequences are discussed The notion of a stable sequence of events generalizes the notion of a mixing sequence of events, introduced in a previous paper of the author. A stable sequence is mixing if its local density is constant almost everywhere.

1. Introduction

Let $[\Omega, \mathscr{A}, P]$ be a probability space in the sense of Kolmogoroff, i.e. let Ω be an arbitrary set whose elements shall be denoted by ω and called elementary events, $\mathscr{A}$ a σ-algebra of subsets of Ω whose elements will be denoted by capital letters A, B etc., and called random events or simply events and $P = P(A)$ a measure, i.e. a non-negative and σ-additive set function defined on $\mathscr{A}$ and normed by the condition $P(\Omega) = 1$; $P(A)$ will be called the probability of the event $A \epsilon \mathscr{A}$.

We shall denote by ϕ the empty set which represents the impossible event, further by $A+B$ the union and by $A.B$ the intersection of the sets A and B. If $A_n (n = 1, 2, \ldots)$ is a sequence of sets, we shall use also the notation $\sum_n A_n$ for the union of the sets A_n. We denote by $A-B$ the set of elements which belong to A but not to B and put $\Omega - A = \bar{A}$. If A and B are arbitrary events such that $P(B) > 0$, we shall denote by $P(A \mid B)$ the conditional probability of the event A with respect to the condition B, i.e. we put

$$P(A \mid B) = \frac{P(AB)}{P(B)}.$$

We shall denote by $a \epsilon A$ that a is an element of the set A and by $A \subset B$ that the set A is a subset of the set B.

As usual a real function $\xi = \xi(\omega)$ defined on Ω is called a random variable if it is measurable with respect to $\mathscr{A}$, that is, if denoting by $\xi^{-1}(I)$ the set of those $\omega \epsilon \Omega$ for which $\xi(\omega) \epsilon I$, then $\xi^{-1}(I)$ belongs to $\mathscr{A}$ if I is an arbitrary interval of the real line.

We denote by $E(\xi)$ the mean value (expectation) of the random variable ξ, i.e. we put $E(\xi) = \int_\Omega \xi dP$.

The infinite sequence of events $A_1, A_2, \ldots, A_n, \ldots$, i.e. of subsets of Ω belonging to $\mathscr{A}$ will be called a *stable* sequence, if the limit

$$\lim_{n \to +\infty} P(A_n B) = Q(B) \qquad \cdots \quad (1.1)$$

exists for every $B \epsilon \mathcal{A}$. We shall show that in this case $Q(B)$ is a bounded measure on $\mathcal{A}$, which is absolutely continuous with respect to the measure P, and thus

$$Q(B) = \int_B \alpha dP \qquad \dots \quad (1.2)$$

for every $B \epsilon \mathcal{A}$ where $\alpha = \alpha(\omega)$ is a measurable function on Ω such that $0 \leqslant \alpha(\omega) \leqslant 1$. We shall call $\alpha(\omega)$ the *local density* of the stable sequence of events $\{A_n\}$.

As well known $\alpha(\omega)$ is not uniquely determined, but if (1.2) holds both with $\alpha = \alpha_1(\omega)$ and $\alpha = \alpha_2(\omega)$ then $\alpha_1(\omega)$ and $\alpha_2(\omega)$ are almost everywhere equal to another.

In the special case when the local density is constant, i.e. $\alpha(\omega) \equiv \alpha$, then $Q(B) = \alpha P(B)$ for every $B \epsilon \mathcal{A}$, i.e. in this case

$$\lim_{n \to +\infty} P(A_n B) = \alpha P(B). \qquad \dots \quad (1.3)$$

Sequences $\{A_n\}$ for which (1.3) holds have been considered already in a previous paper (Rényi, 1958) and have been called strongly *mixing* sequences of events with density α. Thus the notion of a stable sequence of events is a generalization of the notion of a mixing sequence.

The definition of a stable sequence of events can be formulated also in the following equivalent form: The sequence of events $\{A_n\}$ $(n = 1, 2, \dots)$ is called stable if for every event $B \epsilon \mathcal{A}$ such that $P(B) > 0$ the conditional probability $P(A_n | B)$ tends to a limit, i.e.

$$\lim_{n \to +\infty} P(A_n | B) = q(B) \qquad \dots \quad (1.4)$$

exists. Clearly, if $P(B) > 0$ then (1.1) and (1.4) with $q(B) = \frac{Q(B)}{P(B)}$ are equivalent, while if $P(B) = 0$ then (1.1) holds with $Q(B) = 0$ for any sequence $\{A_n\}$.

We shall show that stable sequences of events can be simply characterized in terms of Hilbert space theory. Let H denote the Hilbert space of all random variables ξ, defined on the probability space $[\Omega, \mathcal{A}, P]$, for which $E(\xi^2)$ is finite, the inner product (ξ, η) being defined by $(\xi, \eta) = E(\xi \cdot \eta)$. Let $\alpha_n = \alpha_n(\omega)$ denote the indicator of the set A_n, i.e. $\alpha_n(\omega) = 1$ for $\omega \epsilon A_n$ and $\alpha_n(\omega) = 0$ for $\omega \epsilon \bar{A}_n$. Then the sequence $\{A_n\}$ of events is stable if and only if the sequence α_n converges weakly; the weak limit of the sequence α_n being equal to the local density of the sequence $\{A_n\}$. It follows that the sequence $\{A_n\}$ is mixing if and only if α_n converges weakly to a constant.

We introduce further the notion of a stable sequence of random variables. The sequence of random variables $\xi_n = \xi_n(\omega)$ $(n = 1, 2, \dots)$ will be called *stable* if for any event B with $P(B) > 0$ the conditional distribution of ξ_n with respect to B tends to a limiting distribution, i.e.

$$\lim_{n \to +\infty} P(\xi_n < x | B) = F_B(x) \qquad \dots \quad (1.5)$$

for every x which is a continuity point of the distribution function $F_B(x)$.

Expressed in terms of Hilbert space theory this means that for every bounded and continuous function $g(x)$ the sequence $g(\xi_n)$ converges weakly.

An other equivalent definition of a stable sequence of random variables is the following : the sequence of random variables $\xi_n (n = 1, 2, \ldots)$ is called *stable*, if for every $x \epsilon X$ where X is a set of real numbers which is everywhere dense on the real line, the sequence of events $\xi_n < x$ $(n = 1, 2, \ldots)$ is stable.

Clearly such a sequence $\{\xi_n\}$ of random variables is stable in the sense of (1.5), because if (1.5) holds for x belonging to an everywhere dense set X then it holds for every continuity point x of $F_B(x)$. On the other hand (1.5) implies the stability of the sequence of events $\xi_n < x$ for $x \epsilon X$ where the set X is everywhere dense on the real line.

In the special case when the limiting distribution $F_B(x)$ does not depend on the choice of B we arrive at the notion of a strongly mixing sequence of random variables, introduced previously (Rényi and Révész, 1958).

The aim of the present paper is to study general properties of stable sequences of events and to give criteria for the stability of a sequence of events which are discussed in Section 2; some examples and applications of these notions in probability theory are discussed in Section 3.

2. Some general theorems on stable sequences of events

Let $\alpha_n = \alpha_n(\omega)$ $(n = 1, 2, \ldots)$ be the indicator of the set A_n, i.e. $\alpha_n(\omega) = 1$ if $\omega \epsilon A_n$ and $\alpha_n(\omega) = 0$ if $\omega \epsilon \bar{A}_n$.

Let H denote the Hilbert space of all random variables ξ for which $E(\xi^2)$ exists, the inner product (ξ, η) being defined by $(\xi, \eta) = E(\xi \,.\, \eta)$. We put further $\|\xi\| = (\xi, \xi)^{1/2}$. All definitions and theorems from Hilbert space theory which will be needed in the sequel can be found, e.g., in Szőkefalvi-Nagy (1942).

We prove first the following theorem.

Theorem 1 : *$\{A_n\}$ is a stable sequence of events, i.e. the limit*

$$\lim_{n\to+\infty} P(A_n B) = Q(B) \qquad \ldots \quad (2.1)$$

exists for every $B \epsilon \mathcal{A}$, if and only if $\{\alpha_n\}$ is a weakly convergent sequence of elements of the Hilbert space H, i.e. if for any $\eta \epsilon H$ the limit

$$\lim_{n\to+\infty} (\alpha_n, \eta) = A(\eta) \qquad \ldots \quad (2.2)$$

exists.

Proof of Theorem 1 : Clearly, if β is the indicator of the set B then $(\alpha_n, \beta) = P(A_n B)$. Thus (2.2) reduces to (2.1) if we substitute β instead of η. Thus to prove Theorem 1 it suffices to show that if the limit (2.2) exists whenever η is the indicator of a set B then it exists for every $\eta \epsilon H$. Clearly, if the limit (2.2) exists for every indicator η, it exists also if η is an arbitrary element of H which takes on only a finite number of values. As to every random variable η for which $E(|\eta|) < +\infty$ and to every $\varepsilon > 0$ one can find, by the definition of the Lebesgue integral (Halmos, 1950) a random variable η_1 which takes on only a finite number of values, such that $E(|\eta - \eta_1|) < \varepsilon$, it follows easily that (2.2) holds not only for every $\eta \epsilon H$ but also for every η for which $E(\eta)$ is finite. Thus Theorem 1 is proved.

As a consequence of Theorem 1 we obtain the following theorem.

Theorem 2 : *If* $\{A_n\}$ *is a stable sequence of events, i.e. if*

$$\lim_{n\to+\infty} P(A_nB) = Q(B) \qquad \ldots \quad (2.3)$$

exists for every $B \epsilon \mathcal{A}$, *then* $Q(B)$ *is a measure on* $\mathcal{A}$ *which is absolutely continuous with respect to the measure* P, *and thus can be represented in the form*

$$Q(B) = \int_B \alpha dP \qquad \ldots \quad (2.4)$$

where $\alpha = \alpha(\omega)$ *is a random variable; we have further* $0 \leqslant \alpha \leqslant 1$.

Proof of Theorem 2 : Clearly $A(\eta)$ defined by (2.2) is a bounded linear operation on H, and thus by a well-known theorem [Szőkefalvi-Nagy (1942)] there exists an $\alpha \epsilon H$ such that $A(\eta)$ can be represented in the form $A(\eta) = (\alpha, \eta)$. It is easy to see that $0 \leqslant \alpha \leqslant 1$. It follows that, denoting by β the indicator of the event B, we have

$$Q(B) = (\alpha, \beta) = \int_B \alpha dP. \qquad \ldots \quad (2.5)$$

Thus $Q(B)$ is a measure, which is absolutely continuous with respect to the measure $P(B)$. We shall call $\alpha = \alpha(\omega)$ the *local density* of the stable sequence $\{A_n\}$.

Now we shall prove a criterion of the stability of a sequence of events which is the generalization of a corresponding criterion for mixing sequences, proved by the author (Rényi, 1958).

Theorem 3 : *Let* $\{A_n\}$ $(n = 1, 2, \ldots)$ *be a sequence of events such that the limit*

$$\lim_{n\to+\infty} P(A_n A_k) = Q_k \qquad \ldots \quad (2.6)$$

exists for $k = 1, 2, \ldots$. *Then the sequence* $\{A_n\}$ *is stable, i.e.* (2.1) *holds for every* $B\epsilon\mathcal{A}$.

Proof of Theorem 3 : Let H_1 denote the subspace of H spanned by the sequence $\{\alpha_n\}$ where α_n is the indicator of the event $A_n (n = 1, 2, \ldots)$, i.e. the closure with respect to the distance $\|\xi-\eta\|$ of the set of all finite linear combinations $\sum_{\kappa=1} c_\kappa \alpha_\kappa$ where $c_1, c_2, \ldots, c_n$ are arbitrary real numbers. Let H_2 denote the set of those elements ξ_2 of H which are orthogonal to every $\xi_1 \epsilon H_1$.

According to a well-known theorem (see Szőkefalvi-Nagy, 1942, p. 8) each element ξ of H can be represented in the form $\xi = \xi_1+\xi_2$ where $\xi_1\epsilon H_1$ and $\xi_2\epsilon H_2$. Now we shall prove that if the limit (2.6) exists for $k = 1, 2, \ldots$ then the limit

$$\lim_{n\to+\infty} (\alpha_n, \xi) = A(\xi) \qquad \ldots \quad (2.7)$$

exists for every $\xi\epsilon H$. To prove this it suffices to show that (2.7) exists if $\xi = \xi_1 \epsilon H_1$, because of the above mentioned decomposition of every $\xi\epsilon H$ into the sum of a $\xi_1\epsilon H_1$ and a $\xi_2\epsilon H_2$; as a matter of fact if $\xi = \xi_2\epsilon H_2$ then (2.7) holds with $A(\xi_2) = 0$ while if the limit (2.7) exists for $\xi = \xi_1$ and $\xi = \xi_2$ it clearly exists for $\xi = \xi_1+\xi_2$ also.

Now if ξ_1 is a linear combination of a finite number of the α_k's then clearly the limit (2.7) exists. Let now ξ_1 be an arbitrary element of H_1. Then to every $\varepsilon > 0$ one can find a finite linear combination $\sum_{\kappa=1}^{N} c_\kappa \alpha_\kappa$ such that

$$\| \xi_1 - \sum_{\kappa=1}^{N} c_\kappa \alpha_\kappa \| < \varepsilon. \qquad \ldots \quad (2.8)$$

But (2.8) implies in view of $\|\alpha_n\| \leqslant 1$ and the inequality $|(\xi, \eta)| \leqslant \|\xi\| \cdot \|\eta\|$, that

$$|(\alpha_n, \xi_1) - \sum_{k=1}^{N} c_k(\alpha_n, \alpha_k)| \leqslant \varepsilon. \qquad \cdots (2.9)$$

Thus it follows that

$$|\overline{\lim_{n\to+\infty}} (\alpha_n, \xi_1) - \underline{\lim_{n\to+\infty}} (\alpha_n, \xi_1)| \leqslant 2\varepsilon. \qquad \cdots (2.10)$$

As $\varepsilon > 0$ is arbitrary it follows from (2.10) that the limit (2.7) exists. Thus Theorem 3 is proved.

Let us put (supposing $P(A_k) > 0$ for $k = 1, 2, \ldots$, which is no essential restriction)

$$q_k = \frac{Q_k}{P(A_k)} \; (k = 1, 2, \ldots) \qquad \cdots (2.11)$$

where Q_k is defined by (2.6).

Note that even if all the numbers q_k ($k = 1, 2, \ldots$) are equal, it is not sure that the sequence $\{A_n\}$ is mixing. See for instance examples 1 and 2 of Section 3. This is true, however, in case $A_1 = \Omega$ as it has been shown by the author (Rényi, 1958).

Another way to express this fact is contained in Theorem 4.

Theorem 4 : *Let $\{A_n\}$ be a stable sequence of events such that* $\lim_{n\to+\infty} P(A_n) = q_0$ *and* $\lim_{n\to+\infty} P(A_nA_k) = q_kP(A_k)$ ($k = 1, 2, \ldots$). *Then the sequence $\{A_n\}$ is mixing if and only if the numbers q_k ($k = 0, 1, \ldots$) are all equal to another.*

Proof of Theorem 4 : The necessity of the condition follows immediately from the definition of mixing sequences.

The sufficiency can be proved as follows : Let α denote the local density of the sequence $\{A_n\}$. If $q_k = q \neq 0 (k = 0, 1, \ldots)$ then clearly

$$(\alpha, \alpha_k) = q(1, \alpha_k) \qquad \cdots (2.12)$$

which can be also expressed as follows :

$$(\alpha, \alpha_k) = qP(A_k). \qquad \cdots (2.13)$$

It follows by passing to the limit from (2.12) that

$$(\alpha, \alpha) = q(1, \alpha) \qquad \cdots (2.14)$$

and from (2.13) that

$$(\alpha, \alpha) = q^2. \qquad \cdots (2.15)$$

It follows that

$$q = (1, \alpha) \qquad \cdots (2.16)$$

and therefore that

$$(\alpha, \alpha) = (1, \alpha)^2 \qquad \cdots (2.17)$$

i.e.,

$$\int_\Omega \alpha^2 dP = (\int_\Omega \alpha dP)^2. \qquad \cdots (2.18)$$

This implies that α is constant almost everywhere and thus by (2.15)

$$\alpha \equiv q. \qquad \cdots (2.19)$$

The case $q = 0$ is trivial.

We should like to add the following remark:

In view of Theorem 1, Theorem 3 is a special case of the following.

Theorem : *A bounded sequence $\{\alpha_n\}$ ($n = 1, 2, \ldots$) of elements of a Hilbert space H is weakly convergent if and only if the limits* $\lim_{n\to+\infty} (\alpha_n, \alpha_k)$ *exist for* $k = 1, 2, \ldots$.

The proof of this assertion is exactly the same as that of Theorem 3. This useful theorem, which is due to E. Schmidt (see Schmeidler, 1954) is well known in Hilbert space theory; e.g. the proof in Szőkefalvi-Nagy, (1942, p. 10) of the theorem that every bounded set is weakly compact, is based essentially on this fact.[1]

Let us add that if $\{A_n\}$ is a stable sequence of events and α_n the indicator of A_n, then the sequence α_n converges strongly in H only in the trivial case when the weak limit of α_n is almost everywhere equal either to 1 or to 0, e.g., in the case mentioned in Example 1. As a matter of fact a necessary and sufficient condition of the strong convergence of α_n to α is $\lim\limits_{n\to+\infty} \|\alpha_n\| = \|\alpha\|$. But

$$\lim_{n\to+\infty} \|\alpha_n\|^2 = \lim_{n\to+\infty} (\alpha_n, 1) = \int_\Omega \alpha dP$$

and $\|\alpha\|^2 = \int_\Omega \alpha^2 dP$. Thus α_n is strongly convergent to α if and only if $\int_\Omega \alpha(1-\alpha)dP = 0$ i.e. if $\alpha(1-\alpha) = 0$ almost everywhere.

In view of Theorem 4 and of the well-known theorem of Hilbert space theory according to which every bounded set is weakly compact, the following theorem holds :

Theorem 5 : *Any sequence $\{A_n\}$ of events contains a subsequence which is stable.*

An interesting feature of the stability of a sequence of events is that unlike such properties as independence, equivalence etc., it remains invariant when the underlying measure is replaced by another which is absolutely continuous with respect to the original measure. Moreover the local density of a stable sequence of events remains also unchanged.

Theorem 6 : *Let $[\Omega, \mathcal{A}, P] = \mathcal{S}$ be a probability space and $\{A_n\}$ a stable sequence of events in $\mathcal{S}$. Let P^* be another probability measure on $\mathcal{A}$ which is absolutely continuous with respect to P. Then the sequence $\{A_n\}$ is stable on the probability space $\mathcal{S}^* = [\Omega, \mathcal{A}, P^*]$ also, with the same local density, i.e. if* (2.1) *and* (2.4) *hold, one has also*

$$\lim_{n\to+\infty} P^*(A_n B) = \int_B \alpha dP^*. \quad \ldots \quad (2.20)$$

Proof of Theorem 6 : By supposition

$$P^*(A) = \int_A \rho dP \qquad \text{for } A \epsilon \mathcal{A} \quad \ldots \quad (2.21)$$

where $\rho = \rho(\omega)$ is a nonnegative random variable and $\int_\Omega \rho dP = 1$. It follows that

$$P^*(A_n, B) = \int_B \alpha_n \rho dP = (\alpha_n, \rho\beta) \quad \ldots \quad (2.22)$$

where β is the indicator of the event $B \epsilon \mathcal{A}$. Thus the existence of the limit (2.20) follows evidently from the remark made in course of proving Theorem 1 that the limit (2.2) exists not only if $\eta \epsilon H$ but also under the single condition that $E(\eta)$ exists and that we have in this case also $A(\eta) = (\eta, \alpha)$. Thus it follows that

$$\lim_{n\to+\infty} P^* (A_n B) = (\alpha, \rho\beta) = \int_B \alpha dP^*.$$

[1] Prof. B. Sz. Nagy has kindly called my attention to this proof.

This proves Theorem 6. Let us mention that for mixing sequences a more general result supposing only the semi-continuity of P^* with respect to P has been proved by Sucheston (1962).

We want to make some further general remarks. It is impossible, except in trivial cases that the convergence in (2.1) should be uniform in B for all $B \epsilon \mathcal{A}$. As a matter of fact, putting $B = A_n$ one has

$$P(A_nB) - Q(B) = P(A_n) - Q(A_n) = \int_\Omega (1-\alpha)\alpha_n dP$$

and this difference tends to $\int_\Omega \alpha(1-\alpha)dP$. Thus the convergence in (2.1) can be uniform only if the local density α is almost everywhere equal to 1 or 0 as in the trivial case of Example 1 in Section 3. Nevertheless the convergence in (2.1) may be uniform in B for $B \epsilon \mathcal{B}$ where $\mathcal{B}$ is some proper subset of $\mathcal{A}$ which does not contain the sets A_n themselves or only a finite number of them. For instance, if there exist events B which are independent from all the events A_n, then these may all be contained in $\mathcal{B}$. If B is such an event then clearly the indicator β of B is uncorrelated with the local density α of the sequence $\{A_n\}$.

3. Examples of stable sequences of events

Example 1 : A sequence of identical events $A_n = A (n = 1, 2, \ldots)$ is evidently stable. Note that in this case the local density $\alpha(\omega)$ is equal to 1 for $\omega \epsilon A$ and to 0 for $\omega \epsilon \bar{A}$. Let us mention that the sequence $A, A, \ldots$ is trivially mixing if $P(A) = 0$ or $P(A) = 1$ but not if $0 < P(A) < 1$.

More generally, if A'_n is a mixing sequence of events with density α, and A any event, the sequence $A_n = A'_n A$ is a stable sequence of events, with local density equal to α on A and 0 on $\bar{A}$.

Example 2 : Let $\{A_n\}$ be a sequence of equivalent events (called also symmetrically dependent events) i.e. suppose

$$P(A_{i_1} A_{i_2} \ldots A_{i_k}) = W_k \quad \ldots (3.1)$$

for any choice of the different indices $i_1 < i_2 < \ldots < i_k$ and for $k = 1, 2, \ldots$ where W_k depends on k only, but not on the choice of the indices $i_1, i_2, \ldots, i_k$. Such a sequence $\{A_n\}$ is evidently stable. As a matter of fact, it is sufficient to suppose the independence of W_k from the indices $i_1, i_2, \ldots, i_k$ for $k = 2$ only. This follows clearly from Theorem 3.

Thus sequences of equivalent events are always stable. Note that in view of Theorem 4 a sequence of equivalent events is mixing if and only if

$$W_2 = W_1^2. \quad \ldots (3.2)$$

It is easy to see however that (3.2) is satisfied if and only if the sequence $\{A_n\}$ is a sequence of independent events.

As a matter of fact according to a well-known theorem due to Khintchine (1952) if $\{A_n\}$ is a sequence of equivalent events and W_k is defined by (3.1) then there exists a distribution function $G(x)$ in the interval $[0, 1]$ such that

$$W_k = \int_0^1 x^k dG(x). \quad \ldots (3.3)$$

As a matter of fact this follows from the following theorem of Hausdorff (1923). If a sequence $\{W_k\}$ is monotonic of every order, i.e. putting $W_0 = 1$

$$\sum_{j=1}^{k} (-1)^j \binom{k}{j} W_{n+j} \geqslant 0$$

for all $n \geqslant 0$ and $k \geqslant 0$ then W_k can be represented in the form (3.3). Now clearly $W_2 = W_1^2$ means by (3.3) that

$$\int_0^1 x^2 dG(x) = (\int_0^1 x dG(x))^2$$

which implies evidently that $G(x)$ is the distribution function of a constant c, i.e.

$$G(x) = \begin{cases} 1 & \text{for } x > c \\ 0 & \text{for } x \leqslant c \end{cases}$$

where of course $c = W_1$; but then according to (3.3) $W_k = W_1^k$ that is

$$P(A_{i_1} A_{i_2} \ldots A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j})$$

for every sequence $i_1 < i_2 < \ldots < i_k$ and therefore the events are independent. Thus we have proved the following theorem.

Theorem 7 : *A sequence of equivalent events is always stable, but it is mixing if and only if the events are completely independent.*

Example 3 : Let Ω be the interval (0, 1), $\mathscr{A}$ the set of measurable subset of Ω, and P the Lebesgue-measure. Let the set A_n be defined as the union of the intervals $\left(\frac{k}{n}, \frac{k+\lambda(k/n)}{n}\right)$ $(k = 0, 1, \ldots n-1)$ where $\lambda(x)$ is a continuous function in the interval $[0, 1]$ such that $0 \leqslant \lambda(x) \leqslant 1$. Then clearly the sequence $\{A_n\}$ is stable with local density $\lambda(x)$.

This follows evidently as for any subinterval I of $[0, 1]$ we have

$$P(A_n I) = \frac{1}{n} \sum_{(k/n) \varepsilon I} \lambda\left(\frac{k}{n}\right) + 0\left(\frac{1}{n}\right) \qquad \ldots \quad (3.4)$$

and the first term on the right of (3.4) is a Riemann sum of the integral $\int_I \lambda(x)dx$. Thus

$$\lim_{n \to +\infty} P(A_n I) = \int_I \lambda(x)dx. \qquad \ldots \quad (3.5)$$

It follows easily (e.g. by Theorem 3) that

$$\lim_{n \to +\infty} P(A_n B) = \int_B \lambda(x)dx \qquad \ldots \quad (3.6)$$

for every measurable set B, which proves our assertion.

Example 4 : Let us consider a stationary Markov chain with a finite number of states 1, 2, ..., s. Let ξ_0, the state of the chain at time $t = 0$, have an arbitrary distribution over the set of states. Let ξ_n denote the state of the chain at time $t = n$ $(n = 1, 2, \ldots)$. Let A_n denote the event that the value of ξ_n belongs to a set E where E is a proper subset of the set of states. Let $p_{ij}^{(k)}$ denote the probability of a transition from state i into state j in k steps. Let us suppose that the limits

$$\lim_{k \to +\infty} p_{ij}^{(k)} = \pi_{ij}$$

exist for all i and j; π_{ij} may depend in general on i. It follows that

$$\lim_{n\to+\infty} P(A_nA_k) = \sum_{i=1}^{s} W_i \left[\sum_{j\epsilon E} p_{ij}^{(k)} \left(\sum_{h\epsilon E} \pi_{jh} \right) \right] \quad \ldots \quad (3.7)$$

where W_i is the probability that the chain started at time $t = 0$ in the state i. Thus by Theorem 3 the sequence $\{A_n\}$ is stable. Note that in case π_{ij} does not depend on i, the sequence $\{A_n\}$ is mixing.

Example 5 : Let $S = [\Omega, \mathscr{A}, P]$ be a probability space. Suppose that $\Omega = \sum_{j=1}^{\infty} \Omega_j$; where $\Omega_j \epsilon \mathscr{A}$ and $P(\Omega_j) > 0$ $(j = 1, 2, \ldots)$. Let $S_j = [\Omega_j, \mathscr{A}_j, P_j]$ be the probability space obtained by putting

$$P_j(A) = P(A \mid \Omega_j) \quad \text{for} \quad A \epsilon \mathscr{A}_j, \qquad j = 1, 2, \ldots.$$

where $\mathscr{A}_j$ denotes the set of all $A \epsilon \mathscr{A}$ such that $A \subset \Omega_j$.

Let $\{A_n^{(j)}\}$ be a mixing sequence of sets in the space S_j, with density α_j, and put $A_n = \sum_{j=1}^{\infty} A_n^{(j)} \Omega_j$. Then $\{A_n\}$ is a stable sequence of sets in S, with local density $\alpha(\omega) = \alpha_j$ for $\omega \epsilon \Omega_j$ $(j = 1, 2, \ldots)$.

Clearly, Example 5 covers all cases in which the local density α of a stable sequence of sets has a discrete distribution. As a matter of fact let $[\Omega, \mathscr{A}, P]$ be a probability space and $\{A_n\}$ a stable sequence in this space with local density α where α is a discrete random variable, taking on the different values α_j $(j = 1, 2, \ldots)$ with positive probabilities. Let Ω_j denote the set of those ω for which $\alpha(\omega) = \alpha_j$. Put $P_j(A) = P(A \mid \Omega_j)$ for $j = 1, 2, \ldots$. Then clearly for any $B \epsilon \mathscr{A}$.

$$\lim_{n\to+\infty} P_j(A_n B) = \lim_{n\to+\infty} \frac{P(A_n B\Omega_j)}{P(\Omega_j)} = \frac{Q(B\Omega_j)}{P(\Omega_j)} = \alpha_j P_j(B).$$

Thus the sequence $\{A_n\}$ is mixing in the probability space $[\Omega, \mathscr{A}, P_j]$ with density α_j $(j = 1, 2, \ldots)$.

Remarks : One can generalize Example 5 by splitting the probability space into a non-denumerable instead of a denumerable set of probability spaces, but in this case some care is necessary to avoid measure-theoretical difficulties. However, in this way we arrive at a decomposition of a stable sequence of events into the union of mixing sequences of events, in the most general case. This can be seen as follows. Let $\{A_n\}$ be a stable sequence of sets with local density $\alpha(\omega)$. Then one can define as usual the conditional probability of the event B with respect to a given value of α, which will be denoted by $P_\alpha(B)$. $P_\alpha(B)$ is a random variable such that

$$P(AB) = \int_A P_\alpha(B) dP \quad \text{for every} \quad B \epsilon \mathscr{A}$$

and every $A \epsilon \mathscr{A}_\alpha$ where $\mathscr{A}_\alpha$ is the least σ-algebra on which α is measurable. By other words $P_\alpha(B)$ is the Radon-Nykodim derivative of the set function $P(AB)$ with B fixed with respect to $P(A)$ on the σ-algebra $\mathscr{A}_\alpha$.

As well known, while in case $B_k \epsilon \mathscr{A}$, $B_k B_l = \phi$ for $k \neq l$ the relation

$$P_\alpha\left(\sum_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} P_\alpha(B_k)$$

holds with probability 1, i.e., except for $\omega \epsilon C$ where C is a set such that $P(C) = 0$ nevertheless one cannot say that $P_\alpha(B)$ is with probability 1 a measure, because the set C of

exceptional values of ω may depend on the sequence $\{B_k\}$ and the union of all possible such sets C may have positive measure or even be of measure 1. As however, $P_\alpha(B)$ is not uniquely determined and its value may be changed on a set of measure 0, it is often possible to find a determination of $P_\alpha(B)$ such that it is with probability one a measure. If this is the case it is easy to see that the sequence $\{A_n\}$ is almost surely mixing with respect to the measure $P_\alpha(B)$, with density α.

Such examples can be constructed by means of the theory of measurable decompositions of Lebesgue-spaces, developed by Rochlin (1949). We do not propose to go into details here, but shall return to this question in another paper.

However, we give one example of a stable sequence of random variables constructed by the same principle as applied in the above Example 5 of stable sequence of events.

Example 6 : Let $\{\xi_n\}$ be a mixing sequence of random variables with limiting distribution $F(x)$ and η an arbitrary random variable having a discrete distribution. Let further $g(u, v)$ be a continuous function of two variables. Then the sequence of random variables

$$G_n = g(\xi_n, \eta) \quad (n = 1, 2, \ldots)$$

is strictly stable. As a matter of fact if the values taken on by η with positive probability are denoted by y_k $(k = 1, 2, \ldots)$ and B_k denotes the event $\eta = y_k$ we have

$$\lim_{n \to +\infty} P(G_n < z | B) = \sum_{k=1}^{\infty} P(B_k | B) \int_{g(x, y_k) < z} dF(x). \qquad \ldots \quad (3.8)$$

We may take for instance $g(u, v) = u+v$ in which case we get

$$\lim_{n \to +\infty} P(G_n < z | B) = \sum_{k=1}^{\infty} P(B_k | B) F(z-y_k) \qquad \ldots \quad (3.9)$$

respectively, we may take $g(u, v) = uv$, in which case, supposing that $y_k > 0$ for $k = 1, 2, \ldots$, we obtain

$$\lim_{n \to +\infty} P(G_n < z/B) = \sum_{k=1}^{\infty} P(B_k/B) F\left(\frac{z}{y_k}\right) \qquad \ldots \quad (3.10)$$

for all values of z for which the function on the right hand side of (3.9) respectively (3.10) is continuous.

REFERENCES

HALMOS, P. (1950) : *Measure Theory*, D. Van Nostrand, Princeton, N. J.

HAUSDORFF, F. (1923) : Momentenprobleme für ein endliches Intervall. *Math. Zeitschrift*, **16**, 220–2248.

KHINTCHINE, A. (1952) : O klassah ekvivalentnykh sobytii, *Dokladi, Akad. Nauk*, SSSR. **85**, 713–714.

RÉNYI, A. (1958) : On mixing sequences of events. *Acta Math. Acad. Sci. Hung.*, **9**, 215–228.

RÉNYI, A. and RÉVÉSZ, P. (1958) : On mixing sequences of random variables. *Acta Math. Acad. Sci. Hung.*, **9**, 389–394.

ROCHLIN, V. A. (1949) : On the fundamental notions of measure theory (in Russian). *Matem. Sbornik* **24**, 107–150.

SUCHESTON, L. (1962) : A zero-one property of mixing sequences of events. *Bull. Amer. Math. Soc.*, **68**, 330–332.

SCHMEIDLER, W. (1954) : Lineare operatoren im Hilbertschen Raum Teubuer, Stuttgurt.

SZŐKEFALVI-NAGY, B. (1942) : *Speaktraldarstellung linearer Transformationen des Hilbertschen Raumes*, Springer, Berlin.

Paper received : February, 1963.

DISCRIMINATION OF GAUSSIAN PROCESSES*

By C. RADHAKRISHNA RAO and V. S. VARADARAJAN

Indian Statistical Institute

SUMMARY. Equivalence and orthogonality properties of Gaussian processes are studied in a general context and the results applied to the study of Gaussian measures on real Hilbert spaces. Expressions for the likelihood ratios of two equivalent Gaussian measures on a real separable Hilbert space are derived and the conditions under which their logarithms are quadratic forms are precisely determined.

1. Contributions of Mahalanobis

We briefly refer to Mahalanobis' pioneering work in multivariate analysis, specially because the problem we have considered is closely linked with a distance function known as Mahalanobis D^2 introduced by him in 1925, to study the affinities or divergences between populations. Since then there have been a number of applications of Mahalanobis D^2 in studying the inter-relationships of groups.

D^2 is a function of the characters measured on the individuals of a population and in practice it is important to know the behaviour of D^2 as the number of characters tend to infinity. Since the number of characters studied will always be limited the classification of populations arrived at by using D^2 will be stable only if D^2 converges to a stable value with increase in the number of characters. Mahalanobis (1937) considered this problem and stated some conditions for convergence of D^2 as axioms necessary for successful classification of populations. In a recent paper, Sneath and Sokal (1962) describe a similar problem and lay down hypotheses for successful classification of *Taxa*, which are not very different from the axioms of Mahalanobis given by him twenty five years ago.

An interpretation of these axioms in terms of genetic factors affecting the observable characters was given by Rao (1954). The ideal conditions under which these axioms hold and the modification needed when the characters are affected by environment were also given. It was further shown that the effect of environment is to reduce the distance (dissimilarity) between populations, a conclusion that must be considered seriously by Taxonomists.

2. Introduction

The simplest problem of discrimination consists in assigning an observation x to one of two n-variate normal populations. The discriminant function which provides a dichotomy of the sample space for assigning observations to one or the other of the populations is the likelihood ratio of the densities of an n-dimensional variable with respect to the two populations. The discriminant function is linear in the observations when the two populations differ only in the mean values and quadratic when the dispersion matrices are different. The computations involved in these cases are well known and straightforward.

---

*This paper has been included in *Contributions to Statistics*, presented to Professor P. C. Mahalanobis on the occasion of his 70th. birthday.

However, there often arise in practical work situations where the observations on an individual are of a more general nature, such as the growth of an individual organism measured continuously, contour of an individual's skull, facial profile etc. Each such observation provides a large number of auxiliary observations like the sizes of an organism at various time points and lengths of various diameters of a skull contour, on which finite dimensional techniques could be applied. The computations, however, become unwieldy if the number of auxiliary observations is large. Methods have, therefore, to be developed by which a dichotomy of the sample space of the compound observations can be obtained in an elegant manner. Various questions arise in this connection.

For a fixed number n of auxiliary observations any dichotomy of the sample space will lead to some errors of classification. The errors, however, decrease as n increases. A question then arises as to whether in a given situation the errors $\to 0$ as $n \to \infty$, in which case perfect discrimination between populations is possible, or the errors stabilize at certain values as n increases. An interesting and important problem is to investigate the necessary and sufficient condition under which perfect discrimination is possible on the basis of a compound observation. The second problem is that of utilizing the compound observation in an effective way, when perfect discrimination does not obtain, to decide on the population from which it has arisen. Some answers are provided for both these problems in this paper. References to previous work are given at appropriate places.

3. Mathematical preliminaries

The main result in this section is a limit theorem, which enables us to compute, in the case of infinite dimensional distributions, the Hellinger distance between a pair of probability distributions. The limit theorem is an easy consequence of the well-known Martingale convergence theorem. Both the theorems of this section are known and have been proved by Kraft (1955). We include their statements for the sake of completeness.

We denote by X an abstract space of points x and by $\mathcal{B}$ a σ-algebra of subsets of X. We shall be concerned with measures on $\mathcal{B}$. Since only finite measures are important for our purposes we shall use the term measure to denote only a finite measure. If μ and ν are two measures we say that μ is *absolutely continuous* with respect to ν, $\mu \ll \nu$ in symbols, if $\mu(A) = 0$ whenever $\nu(A) = 0$; μ is said to be *equivalent* to ν, $\mu \equiv \nu$ in symbols, if $\mu \ll \nu$ and $\nu \ll \mu$. If $\mu \ll \nu$, there is a $\mathcal{B}$-measurable function f on X, $\geqslant 0$, uniquely determined ν-almost everywhere, such that $\mu(A) = \int_A f d\nu$ for all $A \epsilon \mathcal{B}$. Any such f is called the *derivative* of μ with respect to ν and is denoted by $d\mu/d\nu$. μ and ν are said to be orthogonal, $\mu \perp \nu$ in symbols, if there is a set $A \epsilon \mathcal{B}$ such that $\mu(A) = \nu(X-A) = 0$. In order that $\mu \perp \nu$ it is sufficient that for each $\epsilon > 0$ there should exist a set B_ϵ with $\mu(B_\epsilon) < \epsilon$ and $\nu(X-B_\epsilon) < \epsilon$; in fact if A_k is the set B_ϵ with $\epsilon = 2^{-k}$ and we write $A = \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k$, then $\mu(A) = \nu(X-A) = 0$.

If μ is a measure on $\mathcal{B}$, f a μ-integrable function on X, and $\mathcal{B}'$ a σ-algebra included in $\mathcal{B}$, the conditional expectation of f given $\mathcal{B}'$ under μ, $E_\mu(f \mid \mathcal{B}')$ in symbols, is any $\mathcal{B}'$-measurable function f' such that $\int_A f' d\mu = \int_A f d\mu$ for all $A \epsilon \mathcal{B}'$. From this it follows at once that if ν is a measure on $\mathcal{B}$ such that $\nu \ll \mu$, and ν', μ' are the respective restrictions of ν, μ to $\mathcal{B}'$, then $d\nu'/d\mu' = E_\mu(d\nu/d\mu \mid \mathcal{B}')$.

Suppose now p and q are two measures on $\mathcal{B}$. Let λ be any measure such that $p \ll \lambda$, $q \ll \lambda$ (such measures always exist; for instance, $\lambda = p+q$). We write $f = dp/d\lambda$, $g = dq/d\lambda$, and define

$$h(p, q) = \int (fg)^{\frac{1}{2}} d\lambda.$$

The function h has many interesting properties. Before discussing some of these we first remark that $h(p, q)$ is independent of the λ used in defining it. In fact let us write $h_\lambda(p, q)$ for the integral $\int (fg)^{\frac{1}{2}} d\lambda$. Suppose λ' is another measure with $p \ll \lambda'$ and $q \ll \lambda'$. If we write $\mu = \lambda + \lambda'$, then $h_\mu(p, q) = \int (dp/d\mu \,.\, dq/d\mu)^{\frac{1}{2}} d\mu = \int (dp/d\lambda . dq/d\lambda)^{\frac{1}{2}}\ d\lambda/d\mu . d\mu = h_\lambda(p, q)$; and, by a similar reasoning, $h_\mu(p, q) = h_{\lambda'}(p, q)$.

We denote by $\mathcal{P}$ the set of all probability measures on $\mathcal{B}$. If $p, q \,\epsilon\, \mathcal{P}$ we shall call $h(p, q)$ the Hellinger distance between p and q. Even though h is not a distance function over $\mathcal{P}$, $(1-h)^{\frac{1}{2}}$ is a distance function, thus providing a partial justification for our nomenclature. The consideration of h goes back to Hellinger (1909). Obviously $h(p, q) \geqslant 0$. We write $H(p, q) = -\log h(p, q)$.

Theorem 3.1 : (i) $0 \leqslant h(p, q) \leqslant 1$ *for all* $p, q \epsilon \mathcal{P}$ (ii) *if* $\mathcal{B}'$ *is a* σ-*algebra* $\subseteq \mathcal{B}$ *and* p', q' *are the restrictions of* p, q *respectively to* $\mathcal{B}'$, *then* $h(p', q') \geqslant h(p, q)$. (iii) $(1-h)^{\frac{1}{2}}$ *is a distance function over* $\mathcal{P}$ (iv) $h(p, q) = 0$ *if and only if* $p \perp q$.

For the proof the reader may refer to Kraft's paper (1955).

Remark : Since $0 \leqslant h(p, q) \leqslant 1$ for all $p, q \epsilon \mathcal{P}$, $0 \leqslant H(p, q) \leqslant \infty$. $H(p, q) = \infty$ if and only if $p \perp q$.

In many applications X is usually a space of functions and consequently computation of h becomes difficult. In such situations it often happens that the measures on $\mathcal{B}$ are built up from measures on certain sub σ-algebras of $\mathcal{B}$. The following theorem examines this set up.

Theorem 3.2 : *Let* $\mathcal{B}_1 \subseteq \mathcal{B}_2 \ldots \subseteq \mathcal{B}_n \subseteq \ldots$ *be an increasing sequence of sub-σ-algebras of* $\mathcal{B}$ *such that* $\mathcal{B}$ *is generated by* $\bigcup_n \mathcal{B}_n$. *Let* p, q *be two probability measures on* $\mathcal{B}$ *and* p_n, q_n *their respective restrictions to* $\mathcal{B}_n$. *Then*

$$h(p_1, q_1) \geqslant h(p_2, q_2) \geqslant \ldots$$

and
$$h(p, q) = \lim_{n \to \infty} h(p_n, q_n).$$

Proof : Let $\lambda = p+q$ and λ_n the restriction of λ to $\mathcal{B}_n$. Let $f = dp/d\lambda$ and $g = dq/d\lambda$. We then know that $dp_n/d\lambda_n = E_\lambda(f|\mathcal{B}_n)$ and $dq_n/d\lambda_n = E_\lambda(g|\mathcal{B}_n)$. Write $f_n = dp_n/d\lambda_n$ and $g_n = dq_n/d\lambda_n$. Clearly $0 \leqslant f, g, f_n, g_n \leqslant 1$. By the Martingale convergence theorem $f_n \rightarrow E(f|\mathcal{B}) = f$ λ-almost everywhere and $g_n \rightarrow g$ λ-almost everywhere. Since $0 \leqslant (f_n g_n)^{\frac{1}{2}} \leqslant 1$ for all n it follows that $h(p_n, q_n) \rightarrow h(p, q)$. The decreasing character of the sequence $h(p_n, q_n)$ follows from Theorem 3.1.

4. Gaussian processes

We shall now apply the results of the preceding section to the study of a pair of Gaussian processes. Suppose $\{\xi_\alpha : \alpha \epsilon T\}$ is a collection of random variables on X such that $\mathcal{B}$ is the smallest σ-algebra of subsets of X relative to which all the ξ_α are measurable. A probability measure p on $\mathcal{B}$ is said to be *Gaussian* relative to the ξ_α if for any k and $\alpha_1, \ldots, \alpha_k \in T$, the joint distribution of $\xi_{\alpha_1} \ldots, \xi_{\alpha_k}$ is a normal distribution.

A typical and extremely useful class of examples is obtained when we take X to be a vector space over the reals and ξ_α to be a collection of linear functionals over X closed under the formation of linear combinations. Defining $\mathcal{B}$ to be the smallest σ-algebra of subsets of X with respect to which all the ξ_α are measurable, we find that a probability measure p on $\mathcal{B}$ is Gaussian if and only if each ξ_α has a normal distribution under p. Two special cases are of great importance in the applications of our theory.

In the first example X is a real separable infinite dimensional Hilbert space and for each $\alpha \epsilon X$, ξ_α is the linear functional $x \rightarrow (\alpha, x)$ of X. $\mathcal{B}$ is then the class of all Borel subsets of X. If p is a Gaussian measure on $\mathcal{B}$ then there is an element $m \in X$ and a linear operator A of X such that $E_p(\xi_\alpha) = (m, \alpha)$ and $\text{cov}_p(\xi_\alpha, \xi_\beta) = (A\alpha, \beta)$ for all $\alpha, \beta \in X$. m is called the *mean* of p and A the *dispersion operator* of p. A is self-adjoint, non-negative definite and has finite trace. m and A uniquely determine p and to each $m \in X$ and self-adjoint, non-negative definite operator A with finite trace there corresponds a p.[cf. Prohorov (1956)]. In this case if $\alpha_1, \alpha_2, \ldots$ is any sequence in X whose linear combinations are dense in X, $\mathcal{B}$ is generated by sets of the form $\xi_{\alpha_n}^{-1}(B)$ (B a Borel set on the line), and a measure on $\mathcal{B}$ is Gaussian if and only if the joint distributions of the ξ_{α_n} are normal.

In the second example X is the space $C[a, b]$ of all continuous functions on $[a, b]$ and $\mathcal{B}$ the class of Borel subsets of X. For each signed measure α on $[a, b]$ ξ_α is the linear functional $x \rightarrow \int_a^b x(t)d\alpha(t)$. A measure p is Gaussian if and only if for all k and all $t_1, \ldots, t_k \epsilon [a, b]$, the joint distribution of $\xi_{t_1}, \ldots, \xi_{t_k}$ is normal, where for any $t \in [a, b]$, ξ_t is the functional $x \rightarrow x(t)$. If p is Gaussian we can define $m(t) = E_p(\xi_t)$ and $K(s, t) = \text{cov}_p(\xi_s, \xi_t)$. $m \in X$ and K is a symmetric function continuous in both variables. m is called the *mean* and K the *covariance function of* p.

Denote by $\tilde{X}$ the Hilbert space of all functions x on $[a, b]$ with $\int_a^b |x(t)|^2 dt < \infty$. There is an obvious inclusion $X \subseteq \tilde{X}$ and it is easily seen that measures on X can be regarded as measures on $\tilde{X}$. Even though the topological structures of X and $\tilde{X}$ are quite different this difference does not play any significant role in our applications. For any measure p on $\mathscr{B}$ we may denote by $\tilde{p}$ the corresponding measure on $\tilde{X}$. The correspondence $p \rightleftarrows \tilde{p}$ preserves absolute continuity and hence equivalence and orthogonality. If p is the Gaussian measure with mean m and covariance function K, the measure $\tilde{p}$ is Gaussian in the Hilbert space $\tilde{X}$. m is still the mean of $\tilde{p}$ while the dispersion operator of $\tilde{p}$ is the (integral) operator $x \rightarrow Ax$ where $(Ax)(t) = \int_a^b K(t, u)x(u)du$ for all t. A classical example is the Wiener measure with mean m and variance parameter c. It is the Gaussian measure on $C[0, 1]$ with mean m and $K(s, t) = c \min(s, t)$ for $s, t \in [0, 1]$; here $c > 0$ is a constant.

It is a consequence of the early work of Cameron and Martin (1944, 1945) that if one considers on the space $C[0, 1]$ the two Wiener measures with means m_1 and m_2 and the same variance parameter $c > 0$, then the two measures are either equivalent or orthogonal and that equivalence is obtained only if $m_1 - m_2$ is sufficiently smooth. On the other hand, if the means are same and the variance parameters are different, then the measures are always orthogonal. These problems were examined by Segal (1958) in a very general context where he obtained necessary and sufficient conditions for equivalence.

Following up these results it was proved by Feldman (1958) and Hajek (1958) that if two probability measures are Gaussian relative to the same set of random variables, then they are either equivalent or orthogonal. Necessary and sufficient conditions for equivalence were also obtained by these authors.

Our concern in this section is also with a pair of probability measure p_1 and p_2 which are Gaussian relative to the same set of random variables. If the two means are m_1, m_2 and the dispersion matrices are Λ_1, Λ_2, we obtain conditions for equivalence in terms $\Lambda_1, \Lambda_2, m_1$ and m_2. In view of the results of Feldman and Hajek it is enough to determine conditions for orthogonality and this is done using the Hellinger distance.

We now introduce the notations which we shall adhere to throughout the rest of this section. X is an abstract set, $\mathscr{B}$ a σ-algebra of subsets of X and $\xi_1, \xi_2, \ldots$ is a sequence of random variables on X such that $\mathscr{B}$ is the smallest σ-algebra of subsets of X-relative to which all the ξ_n are measurable. We denote by $\mathscr{B}_n$ the smallest σ-algebra relative to which $\xi_1, \ldots, \xi_n$ are measurable. A probability measure on $\mathscr{B}$ is called Gaussian if it is Gaussian relative to the $\{\xi_n : n = 1, 2, \ldots\}$. If p is any Gaussian measure we can form

$$m_j(p) = E_p(\xi_j)$$
$$\lambda_{jk}(p) = \text{cov}_p(\xi_j, \xi_k)$$

We denote by $m(p)$ the sequence $(m_1(p), m_2(p), \ldots)$, called the mean of p, and by $\Lambda(p)$ the infinite matrix $(\lambda_{jk}(p))$. For any integer $n \geqslant 1$ we write $\Lambda_n(p)$ for the $n \times n$ matrix $(\lambda_{jk}(p))_{1 \leqslant j, k \leqslant n}$. Clearly $\Lambda_n(p)$ is the dispersion matrix of $(\xi_1, \ldots, \xi_n)$ under p. $\Lambda(p)$ is called the *dispersion matrix* of p. We shall say that p is *non-singular* if $|\Lambda_n(p)| \neq 0$ for any n (for any matrix C we write $|C|$ for its determinant).

A typical example of our considerations is obtained when X is the space of all sequences $x = (x_1, x_2, \ldots)$ of real numbers and for any n, $\xi_n(x) = x_n$. $\mathcal{B}$ is the smallest σ-algebra of subsets of X relative to which all the ξ_n are measurable. If $m = (m_1, m_2, \ldots)$ is any sequence of real numbers and $\Lambda = (\lambda_{jk})$ any infinite matrix such that Λ_n is positive definite for each n, there is a unique Gaussian measure on $\mathcal{B}$ with mean m and dispersion matrix Λ. We shall denote this measure by $p(m, \Lambda)$. $p(m, \Lambda)$ is clearly non-singular.

Suppose now that p_1 and p_2 are two non-singular Gaussian measures on $\mathcal{B}$. We write $m_j = m(p_j) = (m_{j1}, m_{j2}, \ldots)$ and $\Lambda_j = \Lambda(p_j)$ $(j = 1, 2)$. We define

$$d_n = m_{1n} - m_{2n} \qquad n = 1, 2, \ldots$$

$$\delta = (d_1, d_2, \ldots)$$

and

$$\delta_n = (d_1, d_2, \ldots, d_n).$$

We next introduce the matrices

$$\Lambda = \tfrac{1}{2}(\Lambda_1 + \Lambda_2)$$

$$\Lambda_n = \tfrac{1}{2}(\Lambda_{1n} + \Lambda_{2n})$$

We introduce the quantities.

$$\rho_n = 2 \log |\Lambda_n| - \log |\Lambda_{1n}| - \log |\Lambda_{2n}|$$

$$D_n^2 = \delta_n \Lambda_n^{-1} \delta_n^t \text{ (for any matrix } C, C^t \text{ denotes the transpose of } C).$$

D_n^2 is the well-known Mahalanobis D^2 between two distributions in n-space having mean difference δ_n and the same dispersion matrix Λ_n. Finally we write p_{1n} and p_{2n} for the restrictions of p_1 and p_2 to $\mathcal{B}_n$. We recall that $\mathcal{B}_n$ is the smallest σ-algebra with respect to which $\xi_1, \ldots, \xi_n$ are measurable.

Theorem 4.1 : *For each $n \geqslant 1$, we have*

$$H(p_{1n}, p_{2n}) = \frac{1}{4} \rho_n + \frac{1}{8} D_n^2.$$

The sequences $\{\rho_n\}$ and $\{D_n^2\}$ are both non-negative and monotonic non-decreasing ; and if we write $\rho_\infty = \lim_{n \to \infty} \rho_n$ and $D_\infty^2 = \lim_{n \to \infty} D_n^2$, then

$$H(p_1, p_2) = \frac{1}{4} \rho_\infty + \frac{1}{8} D_\infty^2.$$

In order that $p_1 \equiv p_2$ it is necessary and sufficient that $\rho_\infty < \infty$ and $D_\infty^2 < \infty$ i.e., the sequences $\{\rho_n\}$ and $\{D_n^2\}$ be bounded. In other words p_1 and p_2 are equivalent if and only if (i) *the Gaussian measures $p(0, \Lambda_1)$ and $p(0, \Lambda_2)$ are equivalent and* (ii) *the Gaussian measures $p(m_1, \Lambda)$ and $p(m_2, \Lambda)$ are equivalent.*

Proof: If p'_{1n} and p'_{2n} are the distributions (in n-space) of $(\xi_1, \ldots, \xi_n)$ under p_1 and p_2 respectively, it is easy to see that $h(p_{1n}, p_{2n}) = h(p'_{1n}, p'_{2n})$. Using the formulae for the densities of p'_{1n} and p'_{2n} and applying standard techniques of computing multiple integrals we obtain the formula for $H(p_{1n}, p_{2n})$. In view of Theorem 3.1 we know that $H(p_{1n}, p_{2n}) \geqslant 0$ and increases with n.

Substituting in the formula for $H(p'_{1n}, p'_{2n})$ the values $m_{1i} = m_{2i} = 0$ for $i = 1, 2, \ldots, n$ we see that $\frac{1}{4}\rho_n$ is the Hellinger distance between the n-dimensional distributions of $p(0, \Lambda_1)$ and $p(0, \Lambda_2)$. Part (ii) of Theorem 3.1 now enables us to conclude that $\rho_n \geqslant 0$ and increases with n. Similarly $\frac{1}{8}D_n^2$ is the Hellinger distance between the n-dimensional distributions of $p(m_1, \Lambda)$ and $p(m_2, \Lambda)$. Consequently $D_n^2 \geqslant 0$ and increases with n.

Theorem 3.2 now applies to give the result that $p_1 \perp p_2$ if and only if $H(p_{1n}, p_{2n}) \to \infty$ i.e. if and only if either $\rho \ = \infty$ or $D_\infty^2 = \infty$. Since either $p_1 \perp p_2$ or $p_1 \equiv p_2$ we may conclude that $p_1 \equiv p_2$ if and only if $\rho_\infty < \infty$ and $D_\infty^2 < \infty$.

Finally using the same arguments we conclude that $\frac{1}{4}\rho_\infty = H(p(0, \Lambda_1), p(0, \Lambda_2))$ and $\frac{1}{8} D_\infty^2 = H(p(m_1, \Lambda), \ p(m_2, \Lambda))$. This implies immediately that $p_1 \equiv p_2$ if and only if $p(0, \Lambda_1) \equiv p(0, \Lambda_2)$and $p(m_1, \Lambda) \equiv p(m_2, \Lambda)$. This completes the proof of the theorem.

Corollary : *If* $\Lambda_1 = \Lambda_2$, *then* $p_1 \equiv p_2$ *or* $p_1 \perp p_2$ *according as* $\lim\limits_{n \to \infty} \ \delta_n \Lambda_n^{-1} \ \delta_n^t$ $< \infty$ *or* $= \infty$.

Remark : Kraft (1955) also computes $H(p_{1n}, p_{2n})$. The formula which he derives is not however in the same form as ours.

We shall now examine more closely the conditions under which $\rho_\infty < \infty$. Since Λ_{1n} and Λ_{2n} are, for each n, positive definite matrices, there exists a non-singular matrix S_n such that

$$S_n \Lambda_{1n} S_n^t = I_n$$
$$S_n \Lambda_{2n} S_n^t = L_n$$

where I_n is the $n \times n$ unit matrix and L_n is a diagonal matrix. Let $\lambda_{n1}, \lambda_{n2}, \ldots, \lambda_{nn}$ be the entries of L_n. It is known that they are the roots of the equation (in λ)

$$|\Lambda_{2n} - \lambda \Lambda_{1n}| = 0$$

and that they can also be determined as the eigen-values of the positive-definite matrix $\Lambda_{1n}^{-\frac{1}{2}} \Lambda_{2n} \Lambda_{1n}^{-\frac{1}{2}}$. In particular $\lambda_{ni} > 0$ for all n and i.

Theorem 4.2 : *With the above notation,*

$$\rho_n = \sum_{i=1}^{n} (2 \ \log \ (1 + \lambda_{ni})/2 - \log \ \lambda_{ni}).$$

In order that $\rho_\infty < \infty$ *it is necessary and sufficient that there should be positive constants* c, C, M *such that*

(a) $0 < c \leqslant \lambda_{ni} \leqslant C < \infty$ *for all* n *and* i

(b) $\sum_{i=1}^{n} (\lambda_{ni}-1)^2 \leqslant M < \infty$ *for all* n.

Equivalently, $\rho_\infty < \infty$ *if and only if there exists a positive constant* M' *such that*

$$\sum_{i=1}^{n} (\lambda_{ni}-1)^2/\lambda_{ni} \leqslant M' < \infty \text{ for all } n.$$

Proof : From the formula

$$\rho_n = 2 \log |\Lambda_n| - \log|\Lambda_{1n}| - \log|\Lambda_{2n}|$$

it is clear that ρ_n remains unchanged if we replace Λ_{1n} and Λ_{2n} by $S_n \Lambda_{1n} S_n^t$ and $S_n \Lambda_{2n} S_n^t$ respectively. The expression for ρ_n in terms of the λ_{ni} follows at once.

Suppose now there exists a constant $M' > 0$ such that $\sum_{i=1}^{n} (\lambda_{ni}-1)^2/\lambda_{ni} \leqslant M' < \infty$ for all n. Rewriting $2 \log (1+u)/2 - \log u$ as $\log (1+u)^2/4u$ and using the inequality $\log (1+v) < v$ for $v > 0$ we get

$$\rho_n = \sum_{i=1}^{n} \log [1+(\lambda_{ni}-1)^2/4\lambda_{ni}]$$

$$\leqslant \sum_{i=1}^{n} (\lambda_{ni}-1)^2/4\lambda_{ni}$$

$$\leqslant \frac{1}{4} M'.$$

Thus $\rho_\infty < \infty$.

Conversely, let us suppose that $\rho_\infty < \infty$. This means that $\rho_n \leqslant M''$ for all n, $M'' > 0$ being a constant. Since $(1+u)^2 > 4u$, we have, $2 \log (1+\lambda_{ni})/2 - \log \lambda_{ni} > 0$ for all n and i and hence $2 \log (1+\lambda_{ni})/2 - \log \lambda_{ni} \leqslant M''$ for all n and i. Since $2 \log (1+u)/2 - \log u \to \infty$ as $u \to 0$ and as $u \to \infty$ it follows that there are constants $c > 0$ and $C > 0$ such that $0 < c \leqslant \lambda_{ni} \leqslant C < \infty$ for all n and i. But then

$$\begin{aligned}\log \lambda_{ni} - 2 \log (1+\lambda_{ni})/2 &= \log 4\lambda_{ni}/(1+\lambda_{ni})^2 \\ &= \log (1-(\lambda_{ni}-1)^2/(\lambda_{ni}+1)^2) \\ &\leqslant -(\lambda_{ni}-1)^2/(\lambda_{ni}+1)^2 \quad (\text{since } \log (1-t) \leqslant -t \text{ for } 0 < t < 1) \\ &\leqslant -\frac{c}{(C+1)^2} (\lambda_{ni}-1)^2/\lambda_{ni}\end{aligned}$$

so that

$$\rho_n \geqslant \frac{c}{(C+1)^2} \sum_{i=1}^{n} (\lambda_{ni}-1)^2/\lambda_{ni}.$$

Consequently, if $M' = M'' \dfrac{(C+1)^2}{c}$,

$$\sum_{i=1}^{n} (\lambda_{ni}-1)^2/\lambda_{ni} \leqslant M' < \infty$$

for all n. Thus the finiteness of ρ_∞ is equivalent to the boundedness of $\sum_i (\lambda_{ni}-1)^2/\lambda_{ni}$.

Now if $0 < c \leqslant \lambda_{ni} \leqslant C < \infty$ for all n and i and if $\sum_i (\lambda_{ni}-1)^2 \leqslant M < \infty$ for all n, we see that $\sum_i (\lambda_{ni}-1)^2/\lambda_{ni} \leqslant M/c < \infty$ so that we may conclude that $\rho_\infty < \infty$. On the other hand, if $\rho_\infty < \infty$, the analysis in the preceding paragraph shows that there constants c, $C > 0$ such that $0 < c \leqslant \lambda_{ni} \leqslant C < \infty$ for all n and i and that $\sum_i (\lambda_{ni}-1)^2$ remains bounded as n increases. The proof of the theorem is completed.

It follows from Theorem 4.1, that the Gaussian measures p_1 and p_2 are orthogonal whenever $D^2_\infty = \infty$. It is interesting to remark that this result remains true even if p_1 and p_2 are *not* Gaussian. In fact we have

Theorem 4.3 : *Let p_1 and p_2 be two probability measures on $\mathcal{B}$ and let $\xi_1, \xi_2, \ldots$ be a sequence of random variables having finite second moments under both p_1 and p_2. Let*

$$m_{jn} = E_{pj}(\xi_n) \quad (j = 1, 2)$$

$$d_n = m_{1n} - m_{2n}, \ \delta_n = (d_1, \ldots, d_n)$$

and let Λ_1 and Λ_2 be the dispersion matrices and $\Lambda = \frac{1}{2}(\Lambda_1+\Lambda_2)$. If $D^2_n = \delta_n \Lambda_n^{-1} \delta_n^t \to \infty$, then $p_1 \perp p_2$.

Proof: Clearly it is enough to prove that for each $\epsilon > 0$ there is a set B_ε with $p_1(B_\varepsilon)+p_2(X-B_\varepsilon) < \epsilon$. We shall construct, for each n, a set B_n such that $p_1(B_n)+p_2(X-B_n) \leqslant 8/D^2_n$. Since $D^2_n \to \infty$, this will show that $p_1 \perp p_2$.

Consider now, for a given n, $\xi_1, \ldots, \xi_n$. Let S be a non-singular matrix with entries s_{jk} such that $S\,\Lambda_{1n}\,S^t = I$ and $S\,\Lambda_{2n}\,S^t = L$ where I is the $n \times n$ unit matrix and L a diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$. The λ's are all > 0. Define $\eta_j = \sum_k s_{jk}\,\xi_k$. Then $\eta_1, \ldots, \eta_n$ are uncorrelated under both p_1 and p_2. Write now

$$a_j = E_{p_1}(\eta_j), \quad b_j = E_{p_2}(\eta_j), \quad c_j = a_j - b_j$$

and define $c = (c_1, \ldots, c_n)$. We have the equation $\delta_n = c.(S^{-1})^t$ so that

$$\begin{aligned} D^2_n &= \delta_n \Lambda_n^{-1} \delta_n^t \\ &= c.(S^t)^{-1} \Lambda_n^{-1}\, S^{-1}.\, c^t \\ &= c.(S\Lambda_n S^t)^{-1}.\, c^t \\ &= c.(I+L/2)^{-1}.\, c^t \\ &= 2 \sum_{j=1}^{n} c_j^2/(1+\lambda_j). \end{aligned}$$

Let $r_j = c_j/(1+\lambda_j)$ and write $a = \sum_j r_j a_j$ and $b = \sum_j r_j b_j$. We may assume $a \leqslant b$ (since the case where $a \geqslant b$ can be reduced to this by interchanging p_1 and p_2). Let B_n be the set defined by the inequality

$$r_1\eta_1+\ldots+r_n\eta_n > (a+b)/2.$$

Since this is equivalent to $r_1(\eta_1-a_1)+\ldots+r_n(-a_n) > (b-a)/2$, and since $\text{var}_{p_1}(\eta_j) = 1$ for all j, we have, by Chebyshev's inequality,

$$p_1(B_n) \leqslant 4 \sum_j r_j^2/(b-a)^2.$$

Using a similar argument and remembering that $\text{var}_{p_2}(\eta_j) = \lambda_j$ for all j, we get

$$p_2(X-B_n) \leqslant 4 \sum_j r_j^2\lambda_j/(b-a)^2.$$

Since $b-a = \Sigma_j c_j^2/(1+\lambda_j)$, we obtain

$$p_1(B_n)+p_2(X-B_n) \leqslant 4[\Sigma_j c_j^2/(1+\lambda_j)]/[\Sigma_j c_j^2/(1+\lambda_j)]^2 = 8/D_n^2.$$

As observed earlier, this proves that $p_1 \perp p_2$ and finishes the proof of the theorem.

5. Gaussian measures in Hilbert space

We shall examine in this section questions relating to Gaussian measures in a real, separable infinite dimensional Hilbert space X. If A is the dispersion operator of a Gaussian measure with mean m, it is clear that for any $u \in X$ with $(Au, u) = (A^{\frac{1}{2}}u, A^{\frac{1}{2}}u) = 0$ the entire mass of the measure is concentrated on the set $\{x : (x, u) = (m, u)\}$. Since only elementary arguments are needed to take care of such situations, we shall systematically consider only those A for which $A^{\frac{1}{2}}u$ does not vanish unless $u = 0$. Such A we shall call *non-singular*. Since A is non-negative definite, $A^{\frac{1}{2}}u = 0$ if and only if $Au = 0$. We introduce the partial ordering $\leqslant$ in the set of all self-adjoint operators in X; $A \leqslant B$ if $B-A$ is non-negative definite (denoted by $B-A \geqslant 0$). For any operator A we write $R(A)$ for the range of A i.e. $R(A) = \{Au : u \in X\}$.

Let now p_1 and p_2 be Gaussian measures on X with means m_1 and m_2 and dispersion operators Λ_1 and Λ_2, which we shall assume to be non-singular. We write $\Lambda = \frac{1}{2}(\Lambda_1+\Lambda_2)$. It is easily seen that Λ is also non-singular. Let us write $\delta = m_1 - m_2$. We now have the following theorem.

Theorem 5.1 : *In order that $p_1 \equiv p_2$ it is necessary and sufficient that both the following conditions be satisfied :*

(a) $\delta \in R(\Lambda^{\frac{1}{2}})$;

(b) *there exists a bounded self-adjoint operator T such that* (i) $T \geqslant k.I$ *for some constant* $k > 0$ (ii) $tr\ (T-I)^2 < \infty$ (iii) $\Lambda_2 = \Lambda_1^{\frac{1}{2}} T \Lambda_1^{\frac{1}{2}}$.

Proof : We know from Theorem 4.1 that in order that the equivalence $p_1 \equiv p_2$ hold it is necessary and sufficient that (I) the Gaussian measures with means

m_1 and m_2 and dispersion operator Λ are equivalent *and* (II) the Gaussian measures with means 0 and dispersion operators Λ_1 and Λ_2 are equivalent. We shall prove now that (a) is necessary and sufficient for (I) and (b) is necessary and sufficient for (II).

Let P_n be the projection on the spectral subspace X_n of Λ corresponding to the interval $\left[\frac{1}{n}, \infty\right]$ on the real line. Let p_1' and p_2' be the Gaussian measures with means m_1 and m_2 and dispersion operator Λ and let p_{1n}' and p_{2n}' be their projections on X_n. These are equivalent since X_n is finite dimensional and is non-singular. Since they have the same dispersion operator $\Lambda_n = P_n\Lambda P_n$ we have, after a direct computation,

$$H(p_{1n}', p_{1n}') = \frac{1}{8}(\Lambda_n^{-1}\ \delta_n, \delta_n) = \frac{1}{8}\|\Lambda_n^{-\frac{1}{2}}\delta_n\|^2$$

where $\delta_n = P_n\delta$. In other words, in order that the first equivalence holds it is necesary and sufficient that $\|\Lambda_n^{-\frac{1}{2}}\delta_n\|^2$ stay bounded as n increases. Let $\alpha_n \epsilon X_n$ be such that $\Lambda_n^{\frac{1}{2}}\alpha_n = \delta_n$. Since X_n is a spectral subspace of Λ, P_n commutes with $\Lambda^{\frac{1}{2}}$ and hence $\Lambda^{\frac{1}{2}}\,\alpha_n = \delta_n$. We shall show that $\delta \epsilon R(\Lambda^{\frac{1}{2}})$ if and only if $\|\alpha_n\|$ remains bounded as n increases. For, suppose $\|\alpha_n\| \leqslant C < \infty$ for all n. Then there exists an $\alpha \epsilon X$ and a sequence $n_1 < n_2 < \ldots$ such that $(\alpha_{n_k}, u) \rightarrow (\alpha, u)$ for all $u \epsilon X$. Since $\Lambda^{\frac{1}{2}}$ is symmetric, $(\Lambda^{\frac{1}{2}}\beta, u) = (\beta, \Lambda^{\frac{1}{2}}u)$ so that $(\Lambda^{\frac{1}{2}}\alpha_{n_k}, u) \rightarrow (\Lambda^{\frac{1}{2}}\,\alpha, u)$ for all $u \epsilon X$. But $\Lambda^{\frac{1}{2}}\alpha_{n_k} = \delta_{n_k}$ and $\delta_{n_k} \rightarrow \delta$ and $k \rightarrow \infty$. Thus $\Lambda^{\frac{1}{2}}\alpha = \delta$ proving that $\alpha \epsilon R(\Lambda^{\frac{1}{2}})$. On the other hand, if there is an $\alpha \epsilon X$ such that $\Lambda^{\frac{1}{2}}\alpha = \delta$, then the fact that P_n commutes with $\Lambda^{\frac{1}{2}}$ implies that $\Lambda^{\frac{1}{2}}P_n\alpha = \delta_n$ and hence that $\alpha_n = P_n\alpha$. Thus $\alpha_n \rightarrow \alpha$ and hence $\|\alpha_n\|$ remains bounded as n increases.

We now take up the second equivalence and show that for this to hold (b) is necessary and sufficient. Let us denote the Gaussian measures with means 0, and dispersion operators Λ_1 and Λ_2, by q_1 and q_2. Since $\Lambda_1^{-\frac{1}{2}}$ has a dense domain we can find an orthonormal basis $e_1, e_2, \ldots$ for X such that e_n lies in the domain of $\Lambda_1^{-\frac{1}{2}}$ for all n. Let ξ_n be the linear functional $x \rightarrow (x, \Lambda_1^{-\frac{1}{2}}\ e_n)$. Write

$$t_{mn} = (\Lambda_2\Lambda_1^{-\frac{1}{2}}\ e_m,\ \Lambda_1^{-\frac{1}{2}}\ e_n)$$

$$t_{mn}' = t_{mn} - \delta_{mn} \text{ (Kronecker delta)}$$

and T_n be the matrix $(t_{ij})_{1 \leqslant i, j \leqslant n}$. Moreover we have

$$\mathrm{cov}_{p_1}(\xi_m, \xi_n) = \delta_{mn}$$

$$\mathrm{cov}_{p_2}(\xi_m, \xi_n) = t_{mn}.$$

Now the linear combinations of the vectors $\Lambda_1^{-\frac{1}{2}}e_n$ are dense in X and hence we may conclude that the class of Borel sets in X are generated by sets of the form $\xi_n^{-1}(Y)$ (Y a Borel set on the line). Theorems 4.1 and 4.2 now apply and enable us to conclude that the measures q_1 and q_2 are equivalent if and only if (1) there exist constants k and K such that $0 < k \leqslant \lambda_{ni}' \leqslant K$ for all n and i and (2) there exists a

constant K' such that $\sum_i (\lambda'_{ni}-1)^2 \leqslant K' < \infty$ for all n; here $\lambda'_{n1}, \ldots, \lambda'_{nn}$ are the eigenvalues of T_n. These conditions are obviously equivalent to (1') $k \cdot I_n \leqslant T_n \leqslant K.I_n$ for all n and (2') $\text{tr}(T_n-I_n)^2 \leqslant K' < \infty$ for all n, I_n being the $n\times n$ unit matrix. We shall complete the proof by showing that these are equivalent to (b). Suppose there exists a T satisfying (b). We have, $(Te_i, e_j) = (\Lambda_1^{-\frac{1}{2}} T \Lambda_1^{-\frac{1}{2}} \cdot \Lambda_1^{-\frac{1}{2}} e_i, \Lambda_1^{-\frac{1}{2}} e_j) = (\Lambda_2 \Lambda_1^{-\frac{1}{2}} e_i, \Lambda_1^{-\frac{1}{2}} e_j) = t_{ij}$. Hence $(T'^2 e_i, e_i) = (T'e_i, T'e_i) = \sum_j t'^2_{ij}$ where $T' = T-I$. Since $\text{tr}(T'^2) < \infty$ $\sum_i (T'^2 e_i, e_i) < \infty$ and hence $\sum_i \sum_j t'^2_{ij} < \infty$. This shows that $\text{tr}\,(T_n-I_n)^2 = \sum\sum_{i,j\leqslant n} t'^2_{ij}$ remains bounded as n increases, proving (2'). On the other hand, $T \geqslant k.I$ so that $(Tu, u) \geqslant k(u, u)$ for all u. If we consider this for all u in the subspace spanned by $e_1, \ldots, e_n$ we can deduce that $T_n \geqslant k \cdot I_n$. Since T is bounded, $T \leqslant K \cdot I$ for some K and by an argument similar to the one just given we conclude that $T_n \leqslant K \cdot I_n$. This proves (1'). Suppose conversely that (1') and (2') are satisfied. Since $\text{tr}(T_n-I_n)^2 = \sum\sum_{i,j\leqslant n} t'^2_{ij}$, (2') implies that $\sum_i \sum_j t'^2_{ij} < \infty$. Hence there exists a bounded operator T' that $(T'e_i, e_j) = t'_{ij}$ and $\sum_j (T'e_i, T'e_i) = \sum\sum_{i,\,j} t'^2_{ij} < \infty$. Evidently T' is self-adjoint and hence $T'^2 \geqslant 0$ showing that $\text{tr}\, T'^2 < \infty$. Write $T = T'+I$. Then $(Te_i, e_j) = t_{ij}$. Since $k \cdot I_n \leqslant K \cdot I_n$ for all n, $k(u, u) \leqslant (Tu, u) \leqslant K \cdot (u, u)$ for all u which are finite linear combinations of the e_j's. This proves that $k \cdot I \leqslant T \leqslant K.I$. Finally $([\Lambda_1^{\frac{1}{2}} T \Lambda_1^{\frac{1}{2}}] \Lambda_1^{-\frac{1}{2}} e_i, \Lambda_1^{-\frac{1}{2}} e_j) = (Te_i, e_j) = t_{ij} = (\Lambda_2 \Lambda_1^{-\frac{1}{2}} e_i, \Lambda_1^{-\frac{1}{2}} e_j)$, so that if x and y are any two elements in the linear manifold X' generated by the $\Lambda_1^{-\frac{1}{2}} e_j$, $(\Lambda_1^{\frac{1}{2}} T \Lambda_1^{\frac{1}{2}} x, y) = (\Lambda_2 x, y)$. Since X' is dense in X, we can conclude that $\Lambda_2 = \Lambda_1^{\frac{1}{2}} T \Lambda_1^{\frac{1}{2}}$. This proves that (b) is implied by (1') and (2'). As observed earlier this completes the proof of the fact that the Gaussian measures with means 0 and dispersion operators Λ_1 and Λ_2 are equivalent if and only if (b) is satisfied. Taken with the earlier proof that (2) is necessary and sufficient for the equivalence of the Gaussian measures with means m_1 and m_2 and dispersion operator Λ, this proves the entire theorem.

Remarks : (1) It may be noticed that the condition (b) is not symmetric in Λ_1 and Λ_2. However this is only an apparent difficulty. In fact one can show that if (b) is satisfied, there exists an operator S, bounded and self-adjoint, such that (i) $S \geqslant k' \cdot I$ for some constant $k' > 0$ (ii) $\text{tr}(S-I)^2 < \infty$ (iii) $\Lambda_2^{\frac{1}{2}} S \Lambda_2^{\frac{1}{2}} = \Lambda_1$. In fact $Q \cdot = \Lambda_1^{\frac{1}{2}} T^{\frac{1}{2}}$ is a bounded operator and $QQ^* = \Lambda_2$, so that using the polar decomposition of Q and the fact that $\Lambda_2^{\frac{1}{2}}$ has dense range, we deduce that $Q = \Lambda_2^{\frac{1}{2}} U$ where U is unitary. Thus $\Lambda_1^{\frac{1}{2}} = \Lambda_2^{\frac{1}{2}} U T^{-\frac{1}{2}}$ and consequently we have $\Lambda_1 = \Lambda_2^{\frac{1}{2}} U T^{-1} U^{-1} \Lambda_2^{\frac{1}{2}}$. Since U is unitary and $k.I \leqslant T \leqslant K.I$ $S = UT^{-1} U^{-1}$ is a bounded self-adjoint operator with $S \geqslant \frac{1}{K} \cdot I$. It is easy to conclude from $T \geqslant k.I$ and $\text{tr}(T-I)^2 < \infty$ that $\text{tr}(T^{-1}-I)^2 < \infty$ so that $\text{tr}\,(S-I)^2 < \infty$. Obviously $\Lambda_2^{\frac{1}{2}} S \Lambda_2^{\frac{1}{2}} = \Lambda_1$.

(2) It might be noticed that throughout the proof of Theorem 5.1 only superficial use has been made of the fact that Λ_1 and Λ_2 have finite trace. This is not due to accident. The point is that if Λ_1 and Λ_2 do not have finite traces, even though there are no Gaussian measures with these as dispersion operators, there nevertheless exist

Gaussian *weak distributions* (in the sense of Segal (1958)) with dispersion operators Λ_1 and Λ_2 and the conditions (a) and (b) of Theorem 5.1 are necessary and sufficient for the equivalence of the Gaussian weak distributions with means m_1 and m_2 and dispersion operators Λ_1 and Λ_2. The proof of this needs only trivial modifications in the proof of Theorem 5.1. In fact our proof that (b) is necessary and sufficient for the equivalence of the Gaussian distributions with means 0 and dispersion operators Λ_1 and Λ_2 does not use the existence of $\mathrm{tr}(\Lambda_1)$ and $\mathrm{tr}(\Lambda_2)$ and hence goes over without changes. On the other hand, in the proof that (a) is necessary and sufficient for the equivalence of the Gaussian distributions with means m_1 and m_2 and dispersion operator Λ, the only place where compactness of Λ is used is in deducing that the spectral subspace X_n of Λ corresponding to $\left[\frac{1}{n}, \infty\right]$ is finite dimensional which in turn leads to the finiteness of $H(p'_{1n}, p'_{2n})$ and to the formula $H(p'_{1n}, p'_{2n}) = \frac{1}{8}\|\Lambda_n^{-\frac{1}{2}}\,\delta_n\|^2$. If Λ is not compact we may reach the same results by arguing differently. Notice that over X_n, $\Lambda_n \geqslant \frac{1}{n}\cdot I$ so that Λ_n has a *bounded* inverse thereon. It can be easily shown in this case that $H(p'_{1n}, p'_{2n})$ is finite and equals $\frac{1}{8}\,\|\Lambda_n^{-\frac{1}{2}}\,\delta_n\|^2$. The rest of the proof needs no change.

(3) It can be easily verified that condition (b) is equivalent to the conditions of Feldman (1958) in the special context of Theorem 5.1 when m_1 and m_2 are 0. Feldman's methods are however somewhat different from ours.

(4) If $\Lambda_2 = c\Lambda_1$ where $c > 0$ is a constant, then it follows easily from Theorem 5.1 that the Gaussian weak distributions (means arbitrary) with dispersion operators Λ_1 and Λ_2 are orthogonal whenever $c \neq 1$. This generalizes the classical result concerning Wiener measures.

(5) If Λ is the integral operator $x(t) \to \int_0^1 K(t, u)\; x(u)du$ where $K(s, t) = \min(s, t)$ $0 \leqslant s, t \leqslant 1$, then it is a routine computation to obtain the eigen-values and eigen-functions of Λ. It can then be proved easily that an $x \in C[0, 1]$ lies in $R(\Lambda^{\frac{1}{2}})$(Λ is considered in $L_2(0, 1)$) if and only if x is absolutely continuous and $\int_0^1 |x'(t)|^2 dt < \infty$, x' being the derivative of x. In other words, the Wiener measures with the same variance parameter and difference δ in means are equivalent if and only if δ is absolutely continuous and $\delta' \in L_2(0, 1)$.

6. Likelihood ratios

In the preceding sections we have been examining the conditions under which two given measures are equivalent or orthogonal. From the point of view of discrimination the case when equivalence obtains is the nontrivial one and there arises the important question of deriving the expressions for the discriminant functions. Mathematically, if p_1 and p_2 are the Gaussian measures, the problem is that of computing the logarithm L of the likelihood ratio dp_1/dp_2 as a function on the sample space.

We shall study this question in this section. We shall restrict ourselves to a concrete situation, namely when the measures are defined on a real separable infinite dimensional Hilbert space X. Some such restriction seems necessary if one's aim is to exhibit an explicit formula for L. It would be of interest to obtain the form of L in other concrete situations.

In the case when X is finite dimensional, as soon as the dispersion matrices are nonsingular, the Gaussian measures in question are equivalent and it is easy to write down L. Except for an additive constant, L is a linear function if the dispersion matrices are identical while the general case leads to an L which is quadratic in the observations. The situation is somewhat more involved in the infinite dimensional case. The reason for this is that in an infinite dimensional vector space linear and quadratic functions are not well behaved unless one imposes certain topological restrictions on them. Thus, even though it is true that L can be regarded as a quadratic function nearly always, it will not in general be the quadratic form associated with a reasonably well-behaved linear transformation. Theorems 6.1, 6.2 and 6.3, which describe the precise conditions under which L can be associated with well-behaved analytic objects in the underlying Hilbert space and exhibit the expressions for L under these circumstances, are the main results of this section.

It is a general feature of all the known results on likelihood ratios that the computations are made on finite dimensional distributions and the final results obtained by a passage to the limit. Such a method runs into difficulties when the Gaussian measures on the Hilbert space have different dispersion operators mainly because the inverses of the dispersion operators are unbounded in the Hilbert space. So far as we are aware there have been no formulae of significant generality for these likelihood ratios. We shall, in this section, give one such general formula. It is of interest to point out that we do not rely upon the method of finite dimensional approximation but use a more direct argument.

Throughout this section X denotes a real separable infinite dimensional Hilbert space. By a dispersion operator in X we mean a linear operator $A \geqslant 0$ with finite trace; p_1 and p_2 denote Gaussian measures with means m_1, m_2 and dispersion operators Λ_1 and Λ_2 respectively. We write, as usual, $m = \frac{1}{2}(m_1+m_2)$ and $\Lambda = \frac{1}{2}(\Lambda_1+\Lambda_2)$. We shall always assume that Λ_1 and Λ_2 are nonsingular as in Section 5.

We shall first examine the case when $\Lambda_1 = \Lambda_2 = \Lambda$. The results are known (Grenander, 1952) but we give a discussion of this case since it serves to introduce our point of view and prepare the ground for the more difficult case when $\Lambda_1 \neq \Lambda_2$.

Suppose then p_1 and p_2 are Gaussian measures with means m_1, m_2 and dispersion operator Λ. We assume that $p_1 \equiv p_2$. In view of our results in Section 5 this means that $\delta = m_1 - m_2 \epsilon R(\Lambda^{\frac{1}{2}})$ and hence that $\| \Lambda^{-\frac{1}{2}} \delta \|^2 < \infty$. If we write $L = \log (dp_1/dp_2)$ then $L(x) = L'(x-m_2)$ where $L' = \log (dp_1'/dp_2')$, p_1' and p_2' being Gaussian measures with means δ and 0 and dispersion operator Λ. To compute L' we may proceed as follows. Since Λ is a nonsingular dispersion operator we can find an orthonormal

basis $e_1, e_2, \ldots$ for X such that for each n $\Lambda e_n = \lambda_n e_n$, $\lambda_n > 0$. Write $\xi_j(x) = \lambda_j^{-\frac{1}{2}}(x, e_j)$. Then $\xi_1, \xi_2, \ldots$ are independent under both p_1' and p_2' and we have $L'(x) = \lim_{n \to \infty} L_n'(x)$ for almost all x where L_n' is the logarithm of the likelihood ratio of the restrictions of p_1' and p_2' to the smallest σ-algebra with respect to which $\xi_1, \ldots, \xi_n$ are measurable. It is easy to check that

$$L_n'(x) = \sum_{k=1}^{n} \lambda_k^{-\frac{1}{2}}(\Lambda^{-\frac{1}{2}}\delta, e_k)(x, e_k) - \tfrac{1}{2}\sum_{k=1}^{n} (\Lambda^{-\frac{1}{2}}\delta, e_k)^2.$$

The second term converges to $\|\Lambda^{-\frac{1}{2}}\delta\|^2$. The first term consists of independent summands with means 0 and variances $(\Lambda^{-\frac{1}{2}}\delta, e_k)^2$, $(k = 1, 2, \ldots)$ under p'_2, and since $\sum_k (\Lambda^{-\frac{1}{2}}\delta, e_k)^2 < \infty$ the first term converges almost surely. Hence we have the expression $L'(x) = \sum_k (\Lambda^{-\frac{1}{2}}\delta, e_k)\xi_k(x) - \frac{1}{2}\|\Lambda^{-\frac{1}{2}}\delta\|^2$ for L'. Since each ξ_k is a linear function it is reasonable to regard L' as a linear function on X. However L' is defined for almost all x and in general cannot be extended to X as a continuous linear function. We shall say that $L (\equiv \log dp_1/dp_2)$ is *linear* if there exists a continuous linear functional γ on X and a constant c such that $L(x) = \gamma(x) + c$ for almost all x.

Theorem 6.1 : *Let p_1 and p_2 be equivalent Gaussian measures with means m_1 and m_2 and dispersion operator Λ. In order that $L = \log dp_1/dp_2$ be linear it is necessary and sufficient that $\delta \epsilon R(\Lambda)$ where $\delta = m_1 - m_2$. In that case,*

$$L(x) = (x, \Lambda^{-1}\delta) - (m_2, \Lambda^{-1}\delta) - \frac{1}{2}(\delta, \Lambda^{-1}\delta)$$

for almost all x.

Proof : Let us first assume that $\delta \epsilon R(\Lambda)$. Then

$$\begin{aligned}(x, \Lambda^{-1}\delta) &= \sum_k (\Lambda^{-1}\delta, e_k)(x, e_k) \\ &= \sum_k (\delta, \Lambda^{-1} e_k)(x, e_k) = \sum_k \lambda_k^{-1}(\delta, e_k)(x, e_k)\end{aligned}$$

and consequently $L'(x) = (x, \Lambda^{-1}\delta) - \frac{1}{2}\|\Lambda^{-\frac{1}{2}}\delta\|^2$. If we recall that $L(x) = L'(x - m_2)$ and note that $\|\Lambda^{-\frac{1}{2}}\delta\|^2 = (\delta, \Lambda^{-1}\delta)$ we obtain the expression $L(x) = (x, \Lambda^{-1}\delta) - (m_2, \Lambda^{-1}\delta) - \frac{1}{2}(\delta, \Lambda^{-1}\delta)$.

Conversely, let us assume that there exists a continuous linear function γ and a constant c such that $L(x) = \gamma(x) + c$ for almost all x. Then $L'(x) = L(x + m_2) = \gamma(x) + c'$ where c' is another constant. There exists $y \epsilon X$ such that $L'(x) = (x, y) + c'$. Now for any u in X, the joint probability distribution ν under p_2 of ξ and η where $\xi(x) = (x, u)$ and $\eta(x) = (x, y)$ is Gaussian with means zero and dispersion matrix V where

$$V = \begin{pmatrix} (\Lambda u, u) & (\Lambda u, y) \\ (\Lambda u, y) & (\Lambda y, y) \end{pmatrix}$$

and since

$$\int \exp[(x, y) + i(x, u)]dp_2'(x) = \int \exp(\eta + i\xi)d\nu(\xi, \eta)$$

we conclude on the basis of an easy computation that the integral on the left side exists and is equal to $\exp[-\frac{1}{2}(\Lambda u, u)+\frac{1}{2}(\Lambda y, y)+i(\Lambda y, u)]$. Putting $u = 0$ we see that

$$\int \exp[(x, y)]dp_2'(x) = \exp\left[\frac{1}{2}(\Lambda y, y)\right];$$

since $L' = \log dp_1'/dp_2'$, we have

$$\int \exp[L'(x)]dp_2'(x) = 1$$

so that $L'(x) = (x, y)-\frac{1}{2}(\Lambda y, y)$. The equation

$$\int \exp[(x, y)-\frac{1}{2}(\Lambda y, y)+i(x, u)]dp_2'(x) = \exp\left[-\frac{1}{2}(\Lambda u, u)+i(\Lambda y, u)\right]$$

now tells us that p_1' is a Gaussian measure with mean Λy and dispersion operator Λ and consequently that $\delta = \Lambda y$. This proves that $\delta \epsilon R(\Lambda)$. The proof of the theorem is now complete.

We now proceed to the general case when $\Lambda_1 \neq \Lambda_2$. In discussing this case we shall first assume that $m_1 = m_2 = 0$. Let p_1 and p_2 be Gaussian measures with means 0 and dispersion operators Λ_1 and Λ_2 respectively. In view of our results of Section 5 there exists a bounded self-adjoint operator S with $S \geqslant k \cdot I$ for some constant $k > 0$ such that $\Lambda_1 = \Lambda_2^{\frac{1}{2}} S \Lambda_2^{\frac{1}{2}}$ and $\mathrm{tr}(S-I)^2 < \infty$. Let $e_1, e_2, \ldots$ be an orthonormal basis for X such that $\Lambda_2 e_n = \lambda_n e_n$ for all n, $\lambda_n > 0$ being a constant. Write $f_n = \lambda_n^{-\frac{1}{2}} e_n$, $\xi_n(x) = (x, f_n)$. Then $\xi_1, \xi_2, \ldots$ are independent with means 0 and variances 1 under p_2 while they have means 0 and covariances $s_{kl} = (Se_k, e_l)$ under p_1. Since S has a pure point spectrum there is an orthonormal basis $g_1, g_2, \ldots$ of X and constants $s_1, s_2, \ldots > 0$ such that $Sg_n = s_n g_n$ for all n. Write $g_j = \sum_k a_{jk} e_k$ and $\eta_j(x) = \sum_k a_{jk} \xi_k(x)$. Since $\sum_k (a_{jk})^2 < \infty$, η_j is finite with probability one under p_2 and hence under p_1 also. Moreover the orthonormality of $g_1, g_2, \ldots$ implies that $\eta_1, \eta_2, \ldots$ are independent with means 0 and variances 1 under p_2. On the other hand $\mathrm{cov}_{p_1}(\eta_k, \eta_l) = \sum_{r,t} a_{kr} a_{lt} s_{rt} = \sum_{r,t} a_{kr} a_{lt} (Se_r, e_t) = (Sg_k, g_l) = \delta_{kl} s_k$ (Kronecker delta). Consequently $\eta_1, \eta_2, \ldots$ are independent with means 0 and variances $s_1, s_2 \ldots$ under p_1. We may thus conclude that $L(x) = \lim L_n(x)$ for almost all x where

$$L_n(x) = -\frac{1}{2} \sum_{k=1}^{n} \left\{ \left(\frac{1}{s_k} - 1 \right) [\eta_k(x)]^2 + \log s_k \right\}.$$

Since $\mathrm{tr}(S-I)^2 < \infty$, $\sum_k \left(\frac{1}{s_k}-1\right)^2 < \infty$ and hence $\sum_k \log s_k$ converges or diverges with $\sum \left(\frac{1}{s_k}-1\right)$. Assume now that $S-I$ is of trace class i.e. $\sum_k \left|\frac{1}{s_k}-1\right| < \infty$. Then $\sum_k \left|\frac{1}{s_k}-1\right| \cdot [\eta_k(x)]^2$ converges with probability one and hence we have

$$L(x) = -\frac{1}{2} \sum_k \left(\frac{1}{s_k}-1\right) [\eta_k(x)]^2 - \frac{1}{2} \sum \log s_k.$$

If Y is the set of all x such that $\eta_k(x)$ is finite for all k and $\sum_k \left|\frac{1}{s_k}-1\right| [\eta_k(x)]^2 < \infty$, it is clear that Y is a linear manifold of probability one and L is, omitting an additive constant, expressed as an absolutely convergent series of summands each of which is a constant multiple of the square of a linear function. In this sense it is possible to regard L as a quadratic function. Notice that the condition that $(S-I)$ is of trace class, which was imposed during the course of the above discussion is stronger than the condition $\mathrm{tr}(S-I)^2 < \infty$ which is necessary and sufficient for equivalence of p_1 and p_2.

This quadratic function is not in general the quadratic form associated with any linear transformation of X. When X is finite-dimensional, L is, upto an additive constant, the quadratic form associated with the matrix $\Lambda_1^{-1}-\Lambda_2^{-1}$. In the infinite dimensional case Λ_1^{-1} and Λ_2^{-1} are *unbounded* operators and consequently $\Lambda_1^{-1}-\Lambda_2^{-1}$ need not be even defined for a large class of vectors. It is therefore not surprising that one is not always able to exhibit L as a quadratic from. Given a function f on X we shall say that f is a *quadratic form* if there exists a closed, densely defined, symmetric operator A (i.e., $(Ax, y) = (x, Ay)$ for all $x, y \epsilon D(A)$)† such that (i) $p_2(D(A)) = 1$, (ii) $f(x) = (Ax, x)+c$ for some constant c and almost all x. The question that we want to examine now is this: when is L a quadratic form in the sense of this definition?

We proceed to prove a series of lemmas. p_1 and p_2 are equivalent Gaussian measures with means 0 and dispersion operators Λ_1 and Λ_2; $e_1, e_2, \ldots$ is an orthonormal basis for X such that $\Lambda_2 e_n = \lambda_n e_n$ for all n, $\lambda_n > 0$ a constant. Since $\mathrm{tr}(\Lambda_2) < \infty$, for any projection P in X, $\mathrm{tr}\,(\Lambda_2^{\frac{1}{2}} P \Lambda_2^{\frac{1}{2}}) < \infty$. We shall denote it by $\pi(P)$. It is easy to show that $\pi(P) = \mathrm{tr}\,(\Lambda_2 P)$. π is countably additive over mutually orthogonal projections.

Lemma 1: *The following statements on a closed densely defined operator A are equivalent.*

(i) $p_2(D(A)) = 1$.

(ii) *$A\Lambda_2^{\frac{1}{2}} e_n$ is defined for all n and* $\sum_n \|A\Lambda_2^{\frac{1}{2}} e_n\|^2 < \infty$.

(iii) *$M = A\Lambda_2^{\frac{1}{2}}$ is defined everywhere, is a bounded operator and $tr(M^*M) < \infty$. If A is in addition self-adjoint then these are also equivalent to*

(iv) $\int_{-\infty}^{\infty} t^2 da(t) < \infty$

where a is the measure on the real line defined by $a(E) = \pi(P_E)$, $P(E \rightarrow P_E)$ being the spectral measure of A.

Proof: If M is a bounded operator then $\mathrm{tr}\,(M^*M) < \infty$ if and only if $\sum_n (M^*M\, e_n, e_n) = \sum_n \|Me_n\|^2 < \infty$ and consequently (iii)$\rightarrow$(ii). Conversely let us

† $D(A)$ is the domain of definition of A.

assume that (ii) is valid. If x is a finite linear combination, say $x_1e_1+\ldots+x_re_r$ of norm 1, so that $\sum_1^r |x_i|^2 = 1$, then M is defined for x and $\| Mx \|^2 \leqslant (\sum_{i=1}^r |x_i| \, \| Me_i \|)^2 \leqslant \sum_1^r \| M e_i\|^2 \cdot \sum_1^r |x_i|^2 \leqslant C$ where $C = \sum_1^\infty \|Me_n\|^2$. Therefore $\|Mx\| \leqslant C \|x\|$ for all finite linear combinations x of the e_i's. Since A is closed, it follows easily from this that $M = A\Lambda_2^{\frac{1}{2}}$ is defined everywhere and $\|Mx\| \leqslant C \|x\|$ for all x. Since $\sum_n \|Me_n\|^2 < \infty$ (iii) follows immediately. This proves that (ii)$\rightarrow$ (iii).

If A is closed and densely defined there exists a self-adjoint H such that $D(H) = D(A)$ and $\|Hu\| = \|Au\|$ for all $u \epsilon D(A)$; in fact we may take $H = (A^*A)^{\frac{1}{2}}$ (Riesz-Nagy (1952)). This means that for the equivalences (i) $\longleftrightarrow$ (iii) $\longleftrightarrow$ (iv) it may be assumed that A is self-adjoint. We shall do so.

Suppose now (i) is true. We shall prove that (i)$\rightarrow$(iv). For any $x \epsilon X$ let ν_x be the measure $E \rightarrow \|P_E x\|^2$ on the line, P being the spectral measure of A. We know that $D(A) = \{x: \int_{-\infty}^{\infty} t^2 d\nu_x(t) < \infty\}$. Write $D_j = \{t : j \leqslant |t| < j+1\}$ $j = 0, 1, 2, \ldots$ Then $x \epsilon D(A)$ if and only if $\sum_j \int_{D_j} t^2 d\nu_x(t) < \infty$. Since $p_2(D(A)) = 1$ and $\int_{D_j} t^2 d\nu_x(t) \geqslant j^2 \nu_x(D_j)$ it follows that $\sum_j j^2 \nu_x(D_j) < \infty$ for almost all x. Now let $h_{j\mu} (\mu = 1, 2, \ldots)$ be an orthonormal basis for the range of P_{D_j} and let $v_{j\mu}(x) = (P_{D_j} x, h_{j\mu})$. Then $\sum_{j,\mu} (v_{j\mu}(x))^2 = \sum_j j^2 \nu_x(D_j) < \infty$ with probablity one. Since the joint distributions of the $v_{j\mu}$ are all Gaussian and since $E_{p_2}(v_{j\mu}) = 0$ we easily conclude that $\Sigma E_{p_2} v_{j\mu}^2 < \infty$ i.e that $\sum_j j^2 \; E_{p_2} (\phi_j) < \infty$ where E_{p_2} denotes expectation with respect to p_2 and $\phi_j(x) = \nu_x(D_j)$. Since $E_{p_2}(\phi_j) = \pi(P_{D_j}) = \mathrm{tr}(P_{D_j}\Lambda_2) = a(D_j)$, we get $\sum_j j^2 a(D_j) < \infty$. This proves that $\int_{-\infty}^{\infty} t^2 da(t) < \infty$. Conversely from the finiteness of $\int_{-\infty}^{\infty} t^2 da(t)$ we may conclude that $\sum_j (j+1)^2 \, a(D_j) = \sum_j (j+1)^2 \, E_{p_2}(\phi_j) < \infty$ from which the convergence for almost all x of $\sum_j \int_{D_j} t^2 d\nu_x(t) = \int_{-\infty}^{\infty} t^2 d\nu_x(t)$ follows at once. This proves that (iv)$\rightarrow$(i) and completes the proof that (i)$\longleftrightarrow$ (iv).

We now proceed to the equivalence (iv) $\longleftrightarrow$ (iii). If $\|x\| = 1$, $\nu_{\Lambda_2^{\frac{1}{2}}x}(E) = \|P_E \Lambda_2^{\frac{1}{2}} x\|^2 = (\Lambda_2^{\frac{1}{2}} P_E \Lambda_2^{\frac{1}{2}} x, x) \leqslant \mathrm{tr}\,(\Lambda_2^{\frac{1}{2}} P_E \Lambda_2^{\frac{1}{2}}) = \pi(P_E) = a(E)$. Thus, if $\int_{-\infty}^{\infty} t^2 da(t) < \infty$ then $\int_{-\infty}^{\infty} t^2 d\nu_y(t) < \infty$ for $y = \Lambda_2^{\frac{1}{2}} x$, x being arbitrary and of norm 1. This proves that $A\Lambda_2^{\frac{1}{2}}$ is defined everywhere. Moreover, for any $n, a_n(E) = \sum_{j=1}^n \nu_{\Lambda_2^{\frac{1}{2}} e_j}(E) = \sum_1^n (\Lambda_2^{\frac{1}{2}} P_E \, \Lambda_2^{\frac{1}{2}} e_j, e_j) \leqslant \mathrm{tr}(\Lambda_2^{\frac{1}{2}} P_E \Lambda_2^{\frac{1}{2}}) = a(E)$ so that $\sum_1^n \| A \Lambda_2^{\frac{1}{2}} e_j \|^2 = \int_{-\infty}^{\infty} t^2 da_n(t) \leqslant \int_{-\infty}^{\infty} t^2 da(t) < \infty$ for all n. This proves (iii). Finally let us assume that (iii) is valid. Since $a_n(E) \uparrow a(E)$ for all

Borel sets E and $\int_{-\infty}^{\infty} t^2 da_n(t) \leqslant \sum_1^{\infty} \|A\Lambda_2^{\frac{1}{2}} e_n\|^2 < \infty$ for all n, it easily follows that $\int_{-\infty}^{\infty} t^2 d\, a(t) < \infty$. This prove that (iii)$\longleftrightarrow$(iv) and finishes the proof of the lemma.

Lemma 2 : *Let A be a closed, densely defined operator with $p_2(D(A)) = 1$ and let $M = A\Lambda_2^{\frac{1}{2}}$. Then $\Lambda_2^{\frac{1}{2}} M$ is a bounded self-adjoint operator of trace class.*

Proof : Since M is bounded so is $\Lambda_2^{\frac{1}{2}}M$ and since $\Lambda_2^{\frac{1}{2}}M$ is symmetric it is self-adjoint. Further for any $x \epsilon X$, $|(\Lambda_2^{\frac{1}{2}}A\Lambda_2^{\frac{1}{2}}x, x)| = |(A\Lambda_2^{\frac{1}{2}}x, \Lambda_2^{\frac{1}{2}}x)| \leqslant \|\Lambda_2^{\frac{1}{2}}x\| \cdot \|A\Lambda_2^{\frac{1}{2}}x\| \leqslant \frac{1}{2}(\|\Lambda_2^{\frac{1}{2}}x\|^2 + \|A\Lambda_2^{\frac{1}{2}}x\|^2) \leqslant (Vx, x)$ where $V = \frac{1}{2}(\Lambda_2 + M^*M)$. Since this is true for all x and since V is a nonnegative operator of finite trace we may conclude that $\Lambda_2^{\frac{1}{2}}A\Lambda_2^{\frac{1}{2}}$ is of trace class.

Lemma 3 : *Let A be a closed, symmetric, densely defined operator with $p_2(D(A)) = 1$. Suppose further that $((A+\Lambda_2^{-1})x, x) > 0$ for all nonzero $x\epsilon R(\Lambda_2)$. Then $K = \Lambda_2^{\frac{1}{2}}A\Lambda_2^{\frac{1}{2}} + I$ is a bounded self-adjoint operator for which $K \geqslant k.I$ for some $k > 0$ so that K has a bounded inverse. Moreover, $A+\Lambda_2^{-1}$(defined on $R(\Lambda_2)$) is a closed symmetric operator.*

Proof : By Lemma 2, K is a bounded self-adjoint operator. Since $A\Lambda_2^{\frac{1}{2}}$ is bounded, $\Lambda_2^{\frac{1}{2}}A\Lambda_2^{\frac{1}{2}}$ is compact and hence K has a pure point spectrum with its eigen-values converging to 1. For any $x\epsilon R(\Lambda_2)$,$((A+\Lambda_2^{-1})x, x) = (K\,\Lambda_2^{-\frac{1}{2}}x, \Lambda_2^{-\frac{1}{2}}x) \geqslant 0$ so that $K \geqslant 0$. Thus in order to prove that $K \geqslant k.I$ for some $k > 0$ it is enough to prove that 0 is not an eigen-value of K. Suppose for some $y \neq 0$, $Ky = \Lambda_2^{\frac{1}{2}}(A\Lambda_2^{\frac{1}{2}})y + y = 0$. Then $\Lambda_2^{\frac{1}{2}}(A\,\Lambda_2^{\frac{1}{2}}y) = -y$ showing that $y\epsilon R(\Lambda_2^{\frac{1}{2}})$ and $A\Lambda_2^{\frac{1}{2}}y = -\Lambda_2^{-\frac{1}{2}}y$. If $x = \Lambda_2^{\frac{1}{2}}y$, then $x\epsilon R(\Lambda_2)$ $\Lambda_2^{-1}x = \Lambda_2^{-\frac{1}{2}}y$ and $Ax + \Lambda_2^{-1}x = 0$. Since $x \neq 0$ and $((A+\Lambda_2^{-1})x, x) = 0$, we have a contradiction. Finally we claim that $A+\Lambda_2^{-1}$, defined on $R(\Lambda_2)$, is a closed symmetric operator. The symmetry is obvious. Suppose $x_n\epsilon R(\Lambda_2)$ for all n and $x_n \to x$, $Ax_n + \Lambda_2^{-1}x_n \to z$. Write $x_n = \Lambda_2^{\frac{1}{2}}y_n$. Then $\Lambda_2^{\frac{1}{2}}(Ax_n + \Lambda_2^{-1}x_n) = Ky_n \to \Lambda_2^{\frac{1}{2}}z$ and since K^{-1} is bounded, $y_n \to$ a limit, say y. Since $A\Lambda_2^{\frac{1}{2}}$ is bounded, $A\Lambda_2^{\frac{1}{2}}y_n = Ax_n$ tends to a limit and hence $\Lambda_2^{-1}x_n$ tends to a limit. Since Λ_2^{-1} is closed, $x\epsilon D(\Lambda_2^{-1}) = R(\Lambda_2)$. It is then obvious that $Ax_n + \Lambda_2^{-1}x_n \to Ax + \Lambda_2^{-1}x$. The lemma is proved.

Lemma 4 : *Let A be a closed, symmetric, densely defined operator with $p_2(D(A)) = 1$ and let $((A+\Lambda_2^{-1})x, x) > 0$ for all nonzero $x\epsilon R(\Lambda_2)$. Then there exists a unique nonsingular dispersion operator Λ_1' such that $A + \Lambda_2^{-1} = \Lambda_1'^{-1}$.*

Proof : Since $K \geqslant k.I$ we see that $((A+\Lambda_2^{-1})x, x) \geqslant k(\Lambda_2^{-1}x, x)$ for all $x\epsilon R(\Lambda_2)$ and since $\Lambda_2^{-1} \geqslant k'.I$ for $k' > 0$ we see that for any $x\epsilon R(\Lambda_2)$ with $\|x\| = 1$, $\|(A+\Lambda_2^{-1})x\| \geqslant |((A+\Lambda_2^{-1})x, x)| \geqslant kk' > 0$. Since $A+\Lambda_2^{-1}$ is closed, it follows from this that the range of $A+\Lambda_2^{-1}$ is a *closed* linear manifold in X, $\Lambda_1' = (A+\Lambda_2^{-1})^{-1}$ is defined on this closed linear manifold, and is a bounded linear transformation thereon. We now claim that this range is the whole of X. It is enough to prove that it is dense in X. Suppose now for some u, $((A+\Lambda_2^{-1})x, u) = 0$ for all $x\epsilon R(\Lambda_2)$. Then $((A+\Lambda_2^{-1})\Lambda_2^{\frac{1}{2}}y, u) = 0$ for all $y\epsilon R(\Lambda_2^{\frac{1}{2}})$ and hence $(My, u) = -(\Lambda_2^{-\frac{1}{2}}y, u)$ for all $y\epsilon R(\Lambda_2^{\frac{1}{2}})$. Since $M(= A\Lambda_2^{\frac{1}{2}})$

is bounded, this shows that $u \epsilon R(\Lambda_2^{\frac{1}{2}})$ and $M^*u+\Lambda_2^{-\frac{1}{2}}u = 0$ so that $(M^*\Lambda_2^{\frac{1}{2}}+I)\Lambda_2^{-\frac{1}{2}}u = 0$. Since $\Lambda_2^{\frac{1}{2}} M$ is bounded self-adjoint, $M^*\Lambda_2^{\frac{1}{2}}+I = K$ so that $K\Lambda_2^{-\frac{1}{2}}u = 0$. Consequently $\Lambda_2^{-\frac{1}{2}} u = 0$ and hence $u = 0$. This proves that $R(A+\Lambda_2^{-1}))=X$ and hence Λ_1' is a bounded operator on X. As it is the inverse of $A+\Lambda_2^{-1}$, Λ_1' is symmetric, $\geqslant 0$ and is nonsingular. Since $K = \Lambda_2^{\frac{1}{2}}A\Lambda_2^{\frac{1}{2}}+I$ it follows by a straightforward computation that $\Lambda_2^{\frac{1}{2}}K^{-1}\Lambda_2^{\frac{1}{2}} = \Lambda_1'$. Now K^{-1} is bounded and hence $\Lambda_1' = \Lambda_2^{\frac{1}{2}} K^{-1}\Lambda_2^{\frac{1}{2}} \leqslant c\Lambda_2$ for some $c > 0$, showing that $\operatorname{tr}(\Lambda_1') < \infty$. This proves that Λ_1' is a dispersion operator. The proof of the lemma is complete.

Let Q_n be the projection on the subspace spanned by $e_1, \ldots, e_n$ and let $Q_n' = I-Q_n$. For any linear operator F whose domain includes all the vectors e_j we write F_n for the $n \times n$ matrix $((Fe_i, e_j))_{1 \leqslant i, j \leqslant n}$. Many properties of the linear operator can be reduced to an analysis of the asymptotic behaviour of the matrices F_n as $n \to \infty$. For example, if F is a bounded self-adjoint operator such that $F \geqslant a.I$ for some $a > 0$ and $F-I$ is of trace class then $|F_n| > 0$ (where $|\quad|$ denotes determinant) for all n and converges to a finite nonzero limit as $n \to \infty$. This limit is independent of the basis used to compute it. We shall write it as $|F|$ and call it the determinant of F. It is easy to show that $|F^{-1}|$ also exists in the above sense and is equal to $|F|^{-1}$.

With Λ_1' as defined in Lemma 4 form the matrix $C_n = (\Lambda_1'^{-1})_n$ i.e. the matrix whose $i-j$-th element is $(\Lambda_1'^{-1}e_i, e_j)$ For any $y \epsilon X$ we denote by y_n the row vector $((y, e_1), \ldots (y, e_n))$.

Lemma 5 : *With the above notation*

$$\lim_{n \to \infty} y_n C_n^{-1} y_n^t = (\Lambda_1' y, y)$$

for all $y \epsilon X$.

Proof: Fix $y \epsilon X$. We know that two Gaussian measures with mean difference $\Lambda_1' y$ and dispersion operator Λ_1' are equivalent and the D_∞^2 between these two measures is $(\Lambda_1' y, y)$. On the other hand, if $\zeta_j(x) = (x, \Lambda_1'^{-1} e_j)$, the dispersion matrix of $\zeta_1, \ldots, \zeta_n$ is C_n while the row vector of mean differences is $((y, e_1), \ldots (y, e_n))$, so that the D_n^2 between the two distributions of $\zeta_1, \ldots, \zeta_n$ is $y_n C_n^{-1} y_n^t$. It now follows from Theorem 4.1 that $y_n C_n^{-1} y_n^t \to (\Lambda_1' y, y)$. This proves the lemma.

Lemma 6 : *Let A be a closed densely defined symmetric operator with $p_2(D(A)) = 1$. Then,*

$$\lim_{n \to \infty} (Q_n A Q_n y, y) = (Ay, y)$$

for almost all y.

Proof: Since $p_2(D(A)) = 1$, it is clearly enough to prove the lemma with A replaced by any one of its extensions. The Hilbert space X being *real* there is at least one self-adjoint extension of A.* We may therefore assume that A itself is self-adjoint. Moreover any self-adjoint A can be written as $A^+ - A^-$ where A^+ and A^-

*cf. Stone (1932), p. 357.

are $\geqslant 0$, self-adjoint and $D(A^+) \supseteq D(A)$, $D(A^-) \supseteq D(A)$. It is thus enough to prove Lemma 6 when A self-adjoint and $\geqslant 0$.

Let P be the spectral measure of A. For any $\varepsilon > 0$ let $F_{j,\varepsilon}$ be the set $\{t : j\varepsilon \leqslant t < (j+1)\varepsilon\}$ and A_ε the operator $\sum_j j\varepsilon \,.\, P_{F_{j,\varepsilon}}$. Since $\int (P_{F_{j,\varepsilon}} y, y)\, dp_2(y) = \mathrm{tr}(\Lambda_2^{\frac{1}{2}} P_{F_{j,\varepsilon}} \Lambda_2^{\frac{1}{2}})$ it follows that $\Sigma\, j\varepsilon.\ \int (P_{F_{j,\varepsilon}} y, y) dp_2(y) = \mathrm{tr}(\Lambda_2^{\frac{1}{2}} A_\varepsilon A_2^{\frac{1}{2}}) < \infty$ and hence that the function $y \to (A_\varepsilon y, y)$ is integrable with $\int (A_\varepsilon y, y) dp_2(y) = \mathrm{tr}(\Lambda_2^{\frac{1}{2}} A_\varepsilon \Lambda_2^{\frac{1}{2}})$. Now for any ε, $D(A_\varepsilon) = D(A)$ and $|(Ay, y) - (A_\varepsilon y, y)| \leqslant \varepsilon(y, y)$ so that $|\mathrm{tr}(\Lambda_2^{\frac{1}{2}} A_\varepsilon \Lambda_2^{\frac{1}{2}}) - \mathrm{tr}(\Lambda_2^{\frac{1}{2}} A \Lambda_2^{\frac{1}{2}})| \leqslant \sum_n |(\Lambda_2^{\frac{1}{2}}(A - A_\varepsilon)\Lambda_2^{\frac{1}{2}} e_n,\ e_n)| \leqslant \varepsilon \sum_n (\Lambda_2^{\frac{1}{2}} e_n,\ \Lambda_2^{\frac{1}{2}} e_n) = \varepsilon\ \mathrm{tr}(\Lambda_2)$ so that $\int (A_\varepsilon y,\ y) dp_2(y) = \mathrm{tr}(\Lambda_2^{\frac{1}{2}} A \Lambda_2^{\frac{1}{2}}) + 0(\varepsilon)$. Since $(A_\varepsilon y, y) \to (A\, y, y)$ as $\varepsilon \to 0$ we see that $\int (A\, y, y) dp_2(y)$ is finite by using Fatou's lemma. Moreover $|\int (A\, y, y) dp_2(y) - \int (A_\varepsilon\, y, y) dp_2(y)| \leqslant \varepsilon \int (y, y) dp_2(y) = 0(\varepsilon)$. We thus see that the function $y \to (Ay, y)$ is p_2-integrable and

$$\int (Ay, y) dp_2(y) = \mathrm{tr}\, (\Lambda_2^{\frac{1}{2}} A \Lambda_2^{\frac{1}{2}}).$$

Now let $\eta_j(x) = (x, e_j)$. The $\eta_1, \eta_2, \ldots$ are independent under p_2. Let $\mathcal{C}_n$ be the smallest σ-algebra with respect to which $\eta_{n+1}, \eta_{n+2}, \ldots$ are all measurable. Then $\mathcal{C}_1 \supseteq \mathcal{C}_2 \supseteq \ldots$, and by the zero-one law, $\bigcap_n \mathcal{C}_n$ consists only of sets which are of measure 0 or 1. Therefore, by the Martingale convergence theorem, if we have any integrable function g, and write $g_n = E_{p_2}(g \mid \mathcal{C}_n)$, then $g_n(y) \to \int g dp_2$ for almost all y. Take for g the function $y \to (Ay, y)$. For any n, $g(y) = (AQ_n y, Q_n y) + 2(Q_n y,\ A(I - Q_n)y) + ((I - Q_n)y,\ (I - Q_n)y)$. Hence, a straightforward computation yields

$$g_n(y) = \mathrm{tr}\, (\Lambda_{2n}^{\frac{1}{2}} A_n \Lambda_{2n}^{\frac{1}{2}}) + (A(I - Q_n)y, (I - Q_n)y)$$

Since $\mathrm{tr}(\Lambda_{2n}^{\frac{1}{2}} A_n \Lambda_{2n}^{\frac{1}{2}}) = \sum_{i=1}^{n} (\Lambda_2^{\frac{1}{2}} A \Lambda_2^{\frac{1}{2}} e_i, e_i) \to \mathrm{tr}(\Lambda_2^{\frac{1}{2}} A\ \Lambda_2^{\frac{1}{2}}) = \int g dp_2$, it follows that $(A(I - Q_n)y,\ (I - Q_n)y) \to 0$ for almost all y. But $(A(I - Q_n)y,\ (I - Q_n)y) = (Ay,\ y) + (Q_n A Q_n y,\ y) - 2(Ay,\ Q_n y) = (Q_n A Q_n y, y) - (Ay,\ y) + \varepsilon_n(y)$ where $\varepsilon_n(y) \to 0$ as $n \to \infty$ for all y. We therefore conclude that $(Q_n A Q_n y, y) \to (Ay, y)$ for almost all y. Lemma 6 is thus proved.

Lemma 7 : *Let A be a closed, densely defined, symmetric operator with $p_2(D(A)) = 1$ and let $g(y) = exp\left[-\frac{1}{2}(Ay, y)\right]$. Then in order that $\int g dp_2 < \infty$ it is necessary and sufficient that $((A + \Lambda_2^{-1})x, x) > 0$ for all non-zero $x \epsilon R(\Lambda_2)$. In that case $\int g(y) dp_2(y) = |K|^{-\frac{1}{2}}$ where $K = \Lambda_2^{\frac{1}{2}} A \Lambda_2^{\frac{1}{2}} + I$ and $|K|$ denotes the determinant of K. Moreover, for any $u \epsilon X$, $g(y) e^{i(y,u)}$ is integrable and $\int g(y) e^{i(y,u)} dp_2(y) = |K|^{-\frac{1}{2}}$ $exp\left[\frac{1}{2}(\Lambda_1' u, u)\right]$.*

Proof: We shall first examine the conditions for the integrability of g. Suppose that $((A + \Lambda_2^{-1})x, x) > 0$ for all non zero $x \epsilon R(\Lambda_2)$. For any n let $f_n(y) = (Q_n A Q_n y, y)$. By Lemma 6, $\exp\left[-\frac{1}{2} f_n(y)\right] \to \exp\left[-\frac{1}{2}(Ay, y)\right] = g(y)$ for almost

all y. Moreover $\int f_n(y)dp_2(y)$ can be easily checked to be equal to the integral J_n where

$$J_n = (2\pi)^{-n/2}|\Lambda_{2n}|^{-\frac{1}{2}} \int \exp\left[-\frac{1}{2}z(A_n+\Lambda_{2n}^{-1})z^t\right]dz$$

and $z = (z_1, \ldots, z_n)$. The assumption on $A+\Lambda_2^{-1}$ implies that the matrix $A_n+\Lambda_{2n}^{-1}$ is positive definite and hence

$$J_n = |\Lambda_{2n}|^{-\frac{1}{2}}|A_n+\Lambda_{2n}^{-1}|^{-\frac{1}{2}} = |\Lambda_{2n}^{\frac{1}{2}}A_n\Lambda_{2n}^{\frac{1}{2}}+I_n|^{-\frac{1}{2}}$$

so that

$$J_n = |K_n|^{-\frac{1}{2}}.$$

Since $\exp\left[-\frac{1}{2}f_n\right] \geqslant 0$ and converges by Lemma 6 to g almost everywhere, and since J_n tends to $|K|^{-\frac{1}{2}}$, it follows from Fatou's lemma that g is integrable with its integral $\leqslant |K|^{-\frac{1}{2}}$.

Conversely let us assume that g is integrable. Let $\xi_j(x) = (x, e_j)$ and $\mathcal{C}_n$ the smallest σ-algebra with respect to which $\xi_{n+1}, \xi_{n+2}, \ldots$ are all measurable. Let $h(y) = g(y) \exp[i(y, u)]$ and define $g_n = E_{p_2}(g|\mathcal{C}_n)$, $h_n = E_{p_2}(h|\mathcal{C}_n)$. The argument given in Lemma 6 may now be used to prove that $g_n(y) \to E_{p_2}(g)$ and $h_n(y) \to E_{p_2}(h)$ for almost all y. Since g is integrable and since $(Ay, y) = (Az, z)+2(z, Aw) + (Aw, w)$ where $z = Q_n y$, $w = (I-Q_n)y$ it follows that for almost all w, the function

$$z \to \exp\left[-\frac{1}{2}((Az, z)+2(Aw, z))\right]$$

is integrable with respect to the n-dimensional normal distribution having dispersion operator $Q_n\Lambda_2 Q_n$. The integrability for even one w implies that the matrix $C_n = A_n+\Lambda_{2n}^{-1}$ is positive definite. A straightforward computation now yields

$$g_n(y) = |\Lambda_{2n}|^{-\frac{1}{2}}|C_n|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(Ay_n, y_n)+\frac{1}{2}z_n C_n^{-1}z_n^t\right]$$

where $y_n = (I-Q_n)y$, z_n is the row vector whose i-th component is $(Q_n Ay_n, e_i)$. A similar computation on h yields

$$h_n(y) = |\Lambda_{2n}|^{-\frac{1}{2}}|C_n|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}(Ay_n, y_n)+i(y_n, u)+\frac{1}{2}z_n C_n^{-1}z_n^t\right.$$

$$\left.-\frac{1}{2}u_n C_n^{-1}u_n^t - iu_n C_n^{-1}z_n^t\right]$$

where u_n is the row vector with i-th component $(Q_n u, e_i)$. Comparing the formulae for g_n and h_n we get

$$h_n(y) = g_n(y) \exp\left[-\frac{1}{2}u_n C_n^{-1}u_n^t\right] \exp\left[i((y_n, u)-u_n C_n^{-1}z_n^t)\right].$$

Let us write J for $\int g dp_2$ and $J(u)$ for $\int h dp_2$. Since $|h_n(y)| \to |J(u)|$ as $n \to \infty$ for almost all y, we may conclude that $g_n(y) \exp\left[-\frac{1}{2}u_n\ C_n^{-1}u_n^t\right]$ converges to $|J(u)|$ for almost all y. Since $g_n(y) \to J$ for almost all y it follows that $u_n C_n^{-1}u_n^t$ has a limit for each $u \epsilon X$.

There exists a unique bounded self-adjoint operator H_n on X such that $(H_n\, u, u) = u_n\, C_n^{-1} u_n^t$ for all u. Since $\lim_{n \to \infty} (H_n u, u)$ exists for all $u \epsilon X$ it follows from a well-known theorem that $\|H_n\| \leqslant k$ for all n, $k > 0$ being a finite constant. This shows that $0 \leqslant C_n^{-1} \leqslant k.I_n$ for all $n(I_n$ being the $n \times n$ unit matrix) and hence that $C_n \geqslant \frac{1}{k}.I_n$ for all n. We thus deduce that $\Lambda_{2n}^{\frac{1}{2}}\, C_n \Lambda_{2n}^{\frac{1}{2}} \geqslant \frac{1}{k}\Lambda_{2n}$ for all n. If K denotes the bounded operator $\Lambda_2^{\frac{1}{2}} A \Lambda_2^{\frac{1}{2}} + I$, this means that $K_n \geqslant \frac{1}{k}\Lambda_{2n}$ for all n and hence $K \geqslant \frac{1}{k}\Lambda_2$. Since Λ_2 is non-singular, this inequality implies that $Kx = 0$ only when $x = 0$. Since K has a pure point spectrum with non-negative eigen-values which converge to 1, this implies that K is invertible and hence $K \geqslant k'.I$ for some $k' > 0$ from which we may conclude that $((A + \Lambda_2^{-1})x, x) > 0$ for all non-zero $x \epsilon R(\Lambda_2)$. From Lemma 5 we know that $u_n C_n^{-1} u_n^t \to (\Lambda'_1 u, u)$.

We shall now evaluate the limit of $h_n(y)$ for each fixed u. Since $g_n(y) \to J$ and $u_n C_n^{-1} u_n^t \to (\Lambda_1' u, u)$ it follows that $\exp\,[i(u, y_n) - iu_n C_n^{-1} z_n^t]$ has a limit for almost all y. Since $(u_n, y) \to 0$, it follows that $\exp\,(-iu_n C_n^{-1} z_n^t)$ has a limit for almost all y. If we write $\beta_n(y) = -u_n C_n^{-1}\, z_n^t$ then β_n is real linear in y and $\exp[i\beta_n(y)]$ has a limit for almost all y, the limit being independent of y ($= J/|\,J(u)|$ in fact). Such a limit must necessarily be equal to 1 and hence $h_n(y) \to J \exp\Big[-\frac{1}{2}(\Lambda'_1 u, u)\Big]$. Since $K \geqslant k'.I$ and since $K - I$ is of trace class, $|K_n| = |\Lambda_{2n}^{\frac{1}{2}}\, C_n \Lambda_{2n}^{\frac{1}{2}}| = |\Lambda_{2n}|.|C_n| \to |K|$ so that $|\Lambda_{2n}|^{-\frac{1}{2}}|C_n|^{-\frac{1}{2}} \to |K|^{-\frac{1}{2}}$. Since $(Ay_n, y_n) = (Ay, y) + (Q_n A Q_n y, y) - 2(y_n, Ay)$ we may conclude from Lemma 6 that $(Ay_n, y_n) \to 0$ for almost all y. We thus finally see since $z_n C_n^{-1}\, z_n^t \geqslant 0$, that $J \geqslant |K|^{-\frac{1}{2}}$ and hence that $g_n(y) \to |K|^{-\frac{1}{2}}$ and $h_n(y) \to |K|$ $\exp\Big|\Big[-\frac{1}{2}(\Lambda_1' u, u)\Big]$. Lemma 7 follows at once from this.

Remark. It might be of interest to illustrate Lemma 7 with some examples. Let $X = L^2(0, 1)$ and Λ_2 the integral operator with kernel $k(s, t) = \min\,(s, t)$. p_2 then is Wiener measure with variance parameter 1. Let $\lambda > 0$ and $A = -\lambda.I$.

A simple computation shows that the eigen-values $\lambda_1, \lambda_2, \ldots$, of Λ_2 are all simple and $\lambda_n = 4\pi^{-2}(2n-1)^{-2}$. According to Lemma 7, the function $x \to \exp\Big[\frac{1}{2}\lambda(x, x)\Big]$ is integrable if and only if $-\lambda(x, x) + (\Lambda_2^{-1}\, x, x) > 0$ for all non-zero $x \epsilon R(\Lambda_2)$. We claim that this is equivalent to the condition $\lambda < \lambda_n^{-1}$ for all n. If $x = e_n$, then $-\lambda(x, x) + (\Lambda_2^{-1}\, x, x) > 0$ for $x = e_n$ gives $\lambda < \lambda_n$. On the other hand suppose $\lambda < \lambda_n^{-1}$ for all n. Then for any $x = \sum_n x_n e_n \,\epsilon\, R(\Lambda_2)$, $-\lambda(x, x) + (\Lambda_2^{-1}\, x, x) = \Sigma(-\lambda + \lambda_n^{-1})x_n^2 \geqslant 0$ with equality attained only when $x_n = 0$ for all n. Lemma 7 then gives the value of the integral as

$$|-\lambda\Lambda_2 + I|^{-\frac{1}{2}} = \Big[\prod_{n \geqslant 1}\Big[1 - \frac{4\lambda\pi^2}{(2n-1)}\Big]\Big]^{-\frac{1}{2}} = [\sec \lambda^{\frac{1}{2}}]^{\frac{1}{2}},$$

$\lambda < \frac{\pi^2}{4}$ being the condition for integrability. Lemma 7 thus yields the formula

$$\int \exp\left[\frac{1}{2}\lambda \int_0^1 x^2(t)dt\right] dp_2(x) = (\sec \lambda^{\frac{1}{2}})^{\frac{1}{2}} (\lambda^{\frac{1}{2}} < \pi/2)$$

(cf. Gelfand and Yaglom (1960) and Silov (1963)).

As another example let us consider the function

$$f_p : x \rightarrow \exp\left[\frac{\lambda}{2} \int_0^1 p(t)x^2(t)dt\right]$$

where $\lambda > 0$ and p is a continuous non-negative function on [0, 1]. Let $m(\Delta) = \inf_{t \in \Delta} p(t)$ for any closed interval $\Delta \subseteq [0, 1]$ and let us write $\sigma_p = \sup_{\Delta} (|\Delta|^2 \, m(\Delta))$($|\Delta|$ denotes the length of Δ). Lemma 7 can then be used to prove that f_p is integrable if and only if the condition

$$\lambda < \frac{\pi^2}{4\sigma_p}$$

is satisfied. When $p(t) = 1, \sigma_p = 1$ leading to the first example. When $p(t) \equiv t^2$ the condition becomes $\lambda < 4\pi^2$.

We are now in a position to formulate and prove our second main theorem of this section.

Theorem 6.2 : *Let p_1 and p_2 be two equivalent Gaussian measures with means 0 and (non-singular) dispersion operators Λ_1 and Λ_2. Let $L = log\ (dp_1/dp_2)$. In order that L be a quadratic form it is necessary and sufficient that $R(\Lambda_1) = R(\Lambda_2)$ and that $(\Lambda_1^{-1} - \Lambda_2^{-1})\Lambda_2^{\frac{1}{2}}$ has a bounded extension M to the whole of X with $tr(M^*M) < \infty$. In this case the closure A of $\Lambda_1^{-1} - \Lambda_2^{-1}$ exists, is a closed symmetric operator with $p_2\ (D(A)) = 1$, and for almost all x*

$$L(x) = -\frac{1}{2}(Ax, x) - \frac{1}{2}\ log\ |S|$$

where S is the operator satisfying the equation $\Lambda_1 = \Lambda_2^{\frac{1}{2}} S \Lambda_2^{\frac{1}{2}}$.

Proof : Suppose first that the conditions on Λ_1 and Λ_2, specified in the theorem, are satisfied. $\Lambda_1^{-1} - \Lambda_2^{-1}$ is then densely defined and since it is symmetric, its closure exists. Let A denote this closure. The finiteness of $\mathrm{tr}(M^*M)$ implies that $\sum_n \| A\Lambda_2^{\frac{1}{2}} e_n \|^2 < \infty$, and hence by Lemma 1, $p_2(D(A)) = 1$. Since $A + \Lambda_2^{-1} = \Lambda_1^{-1}$ on $R(\Lambda_2)$, the conditions of Lemma 7 are satisfied with $\Lambda_1' = \Lambda_1$. Consequently $\exp\left[-\frac{1}{2}(Ax, x)\right]$ is an integrable function with

$$\int \exp\left[-\frac{1}{2}(Ax, x) + i(x, u)\right] dp_2(x) = |K|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\ (\Lambda_1 u, u)\right]$$

for all u. Now $K = \Lambda_2^{\frac{1}{2}} A \Lambda_2^{\frac{1}{2}} + I$ and hence $Ke_n = \lambda_n(\Lambda_1^{-1} - \Lambda_2^{-1})e_n + e_n = \lambda_n \Lambda_1^{-1} e_n = S^{-1} e_n$ for all n. This implies that $K = S^{-1}$ and hence that $|K|^{-\frac{1}{2}} = |S|^{\frac{1}{2}}$. We thus conclude that

$$|S|^{-\frac{1}{2}} \int \exp\left[-\frac{1}{2}(Ax, x) + i(x, u)\right] dp_2(x) = \exp\left[-\frac{1}{2}(\Lambda_1 u, u)\right].$$

This shows that dp_1/dp_2 is the function $x \to |S|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(Ax, x)\right]$ and leads to the result that

$$L(x) = -\frac{1}{2}(Ax, x) - \frac{1}{2}\log\ |S|.$$

Conversely, let us assume that L is a quadratic form. Then there exists a closed, symmetric, densely defined operator A with $p_2(D(A)) = 1$ such that $L(x) = -\frac{1}{2}(Ax, x) + c$ for almost all x, c being a constant. This implies that the function $x \to \exp\left[-\frac{1}{2}(Ax, x)\right]$ is integrable with respect to p_2. From Lemma 1 we infer that $M = A\Lambda_2^{\frac{1}{2}}$ is everywhere defined and $\operatorname{tr}(M^* M) < \infty$. Moreover $R(\Lambda_2) \subseteq D(A)$ and hence $A + \Lambda_2^{-1}$ is defined on $R(\Lambda_2)$. Lemma 7 implies that $A + \Lambda_2'^{-1} = \Lambda_1^{-1}$ where Λ_1' is a non-singular dispersion operator. Evidently $R(\Lambda_1') = R(\Lambda_2)$. Theorem 6.2 will be completely proved if we show that $\Lambda_1' = \Lambda_1$. From Lemma 7, we have, for all $u \in X$,

$$\int \exp\left[-\frac{1}{2}(Ax, x) + i(x, u)\right] dp_2(x) = |S|^{\frac{1}{2}} \exp\left[-\frac{1}{2}(\Lambda_1' u, u)\right].$$

Since $L(x) = -\frac{1}{2}(Ax, x) + c$ and since

$$\int \exp\left[-\frac{1}{2}(Ax, x)\right] dp_2(x) = |S|^{\frac{1}{2}},$$

we conclude from the equation $\int \exp L(x) dp_2(x) = 1$ that

$$c = -\frac{1}{2}\log |S|$$

and hence that

$$\int \exp\left[L(x) + i(x, u)\right] dp_2(x) = \exp\left[-\tfrac{1}{2}(\Lambda_1' u, u)\right].$$

Since $L = \log dp_1/dp_2$, the integral on the left is equal to $\exp\left[-\frac{1}{2}(\Lambda_1 u, u)\right]$. This proves that $\Lambda_1 = \Lambda_1'$ and finishes the proof of Theorem 6.2.

Corollary : *In order that L be almost everywhere equal to $c+q$ where c is a constant and q the quadratic form associated with a bounded self-adjoint operator it is necessary and sufficient that $R(\Lambda_1) = R(\Lambda_2)$ and $\Lambda_1^{-1} - \Lambda_2^{-1}$ be a bounded operator on this linear manifold. If A denotes the bounded extension of $\Lambda_1^{-1} - \Lambda_2^{-1}$ to X, then A is a bounded self-adjoint operator and*

$$L(x) = -\frac{1}{2}(Ax, x) - \frac{1}{2}\log |S|.$$

Remarks : (1) It may be noticed that the conclusion that L is a quadratic form is symmetric between p_1 and p_2 while the conditions derived on Λ_1 and Λ_2 do not exhibit this symmetry. But this is only an apparent difficulty. Suppose in fact that $R(\Lambda_1) = R(\Lambda_2)$ and $\Lambda_1^{-1}-\Lambda_2^{-1}$ has the closure A. We claim that $N = (-A)$. $\Lambda_1^{\frac{1}{2}}$ is an everywhere defined bounded operator with $\mathrm{tr}(N^*N) < \infty$. Since $\Lambda_1 = \Lambda_2^{\frac{1}{2}}S\Lambda_2^{\frac{1}{2}}$, $\Lambda_1^{\frac{1}{2}} = \Lambda_2^{\frac{1}{2}}S^{\frac{1}{2}}U$ where U is an unitary operator and hence $N = -A\Lambda_2^{\frac{1}{2}}S^{\frac{1}{2}}U$ which is everywhere defined. Moreover $N = -MS^{\frac{1}{2}}U$ so that $N^*N = U^{-1}S^{\frac{1}{2}}M^*MS^{\frac{1}{2}}U$. To prove that $\mathrm{tr}\,(N^*N) < \infty$ it suffices to prove that $\mathrm{tr}\,(S^{\frac{1}{2}}M^*MS^{\frac{1}{2}}) < \infty$ and for this it is enough to find an orthonormal basis $f_1, f_2, \ldots,$ of X such that $\sum\limits_n \| S^{\frac{1}{2}}M^*MS^{\frac{1}{2}}f_n\|^2 < \infty$. Since $\mathrm{tr}(S-I)^2 < \infty$, S has a pure point spectrum and hence we can choose an orthonormal basis $f_1, f_2, \ldots$ for X such that $Sf_n = s_nf_n$ for all n, $s_n > 0$ a constant. Since $\sum\limits_n(s_n-1)^2 < \infty$, $s_n \to 1$ as $n \to \infty$. Consequently $\sum\limits_n \| S^{\frac{1}{2}} M^*MS^{\frac{1}{2}}f_n \|^2 \leqslant \| S \| \sum\limits_n \| M^*MS^{\frac{1}{2}}f_n\|^2 \leqslant \|S\| \sum s_n \| M^*Mf_n\|^2 < \infty$ since $s_n \to 1$ and $\sum\limits_n \| M^*Mf_n\|^2 = \mathrm{tr}\,(M^*M) < \infty$.

(2) It might be interesting to notice that even the assumption that $S-I$ is of trace class need not imply the conditions of Theorem 6.2. We shall now describe an example where $S-I$ is of trace class and yet $R(\Lambda_1)$ does not coincide with $R(\Lambda_2)$. Let $\lambda_1, \lambda_2, \ldots > 0$ be chosen so that $\sum\limits_n \lambda_n \leqslant \frac{1}{16}$. Let $e_1, e_2, \ldots$ be an orthonormal basis for X and let $\Lambda_2 e_n = \lambda_n e_n$ for all n. We define a bounded operator S by setting

$$Se_j = \begin{cases} (1+2\lambda_1^{\frac{1}{2}})e_1 + \lambda_2^{\frac{1}{2}}e_2 + \ldots & \text{if } j = 1 \\ \lambda_j^{\frac{1}{2}}e_1 + e_j & \text{if } j > 1. \end{cases}$$

If $x = x_1e_1 + \ldots + x_re_r$ is a vector of norm 1, then

$$(Sx, x) = (x_1^2 + \ldots + x_r^2) + 2x_1(x_1\lambda_1^{\frac{1}{2}} + \ldots + x_r\lambda_r^{\frac{1}{2}}) = 1 + 2x_1(x_1\lambda_1^{\frac{1}{2}} + \ldots + x_r\lambda_r^{\frac{1}{2}})$$

and since $$|x_1(x_1\lambda_1^{\frac{1}{2}} + \ldots + x_r\lambda_r^{\frac{1}{2}})\,| \leqslant |x_1| \sum_{j=1}^{r} |x_j|\,\lambda_j^{\frac{1}{2}} \leqslant (\sum_1^{\infty} \lambda_j)^{\frac{1}{2}} \leqslant 1/4$$

we see that $(Sx, x) \geqslant \frac{1}{2}$ for all such x. This shows that S is a bounded self-adjoint operator with $S \geqslant \frac{1}{2}.I$. Moreover,

$$(S-I)e_j = \begin{cases} 2\lambda_1^{\frac{1}{2}}e_1 + \lambda_2^{\frac{1}{2}}e_2 + \ldots & \text{if } j = 1 \\ \lambda_j^{\frac{1}{2}}\, e_1 & \text{if } j > 1 \end{cases}$$

from which we may conclude, that $S-I$ is of trace class. If we now write

$$\Lambda_1 e_j = \begin{cases} \lambda_1(1+2\lambda_1^{\frac{1}{2}})e_1 + \lambda_1^{\frac{1}{2}} \sum\limits_{k \geqslant 2} \lambda_k e_k & \text{if } j = 1 \\ \lambda_1^{\frac{1}{2}}\,\lambda_j e_1 + \lambda_j e_j & \text{if } j < 1 \end{cases}$$

then $\Lambda_1 = \Lambda_2^{\frac{1}{2}}S\Lambda_2^{\frac{1}{2}}$ is a dispersion operator which is non-singular and the Gaussian measures with means 0 and dispersion operators Λ_1 and Λ_2 are equivalent. In this case $R(\Lambda_1) \neq R(\Lambda_2)$. In fact $e_1 \notin R(\Lambda_1^{-})$. For, suppose $a_1, a_2, \ldots$ are constants with

$\Sigma a_j^2 < \infty$ such that $\Lambda_1(a_1e_1+\ldots) = e_1$. Using the definition of Λ_1 we get the equations

$$\lambda_1^{\frac{1}{2}}\lambda_k a_1 + \lambda_k a_k = 0 \quad (k > 1).$$

The above set of equations imply that $a_k = -\lambda_1^{\frac{1}{2}} a_1$ for all k and hence that $a_k = 0$ for all k. This is a contradiction. In terms of matrices, Λ_2 is a diagonal matrix with diagonal entries $\lambda_1, \lambda_2 \ldots$ while Λ_1 is the matrix (a_{ij}) where $a_{11} = \lambda_1 (1+2\lambda_1^{\frac{1}{2}})$, $a_{1j} = a_{j1} = \lambda_1^{\frac{1}{2}} \lambda_j \, (j > 1)$ and $a_{ij} = \lambda_j \, \delta_{ij} \, (i, j \geqslant 2)$.

(3) All these considerations simplify considerably if Λ_1 and Λ_2 *commute.* In this case we can choose an orthonormal basis $e_1, e_2, \ldots$ for X such that $\Lambda_1 e_n = \mu_n e_n$, $\Lambda_2 e_n = \lambda_n e_n$ for all n, $\lambda_n > 0$ and $\mu_n > 0$ being constants. The condition for equivalence is now $\Sigma \left(\frac{\mu_n}{\lambda_n} - 1 \right)^2 < \infty$. $Se_n = \left(\frac{\mu_n}{\lambda_n} \right) e_n$ for all n so that $S-I$ is of trace class if and only if $\sum_n \left| \frac{\mu_n}{\lambda_n} - 1 \right| < \infty$. Since $\frac{\mu_n}{\lambda_n} \to 1$ it follows easily that $R(\Lambda_1) = R(\Lambda_2)$ and that the closure of $\Lambda_1 - \Lambda_2$ is the unique self-adjoint operator A for which $Ae_n = \left(\frac{1}{\mu_n} - \frac{1}{\lambda_n} \right) e_n$ for all $n.D(A)$ consists of all $x \epsilon X$ for which $\sum_u \left(\frac{1}{\mu_n} - \frac{1}{\mu_n} \right)^2 (x, e_n)^2 < \infty$ and this has probability one if and only if $\sum_n \left(\frac{1}{\mu_n} - \frac{1}{\lambda_n} \right)^2 \lambda_n < \infty$ which is also the condition for L to be a quadratic form. This however need not always happen. If $\lambda_n = 1/n^2$ and $\mu_n = 1/n^2 + n^{\frac{1}{2}}$ then $\sum_n \left| \frac{\mu_n}{\lambda_n} - 1 \right| < \infty$ but $\sum_n \left(\frac{1}{\mu_n} - \frac{1}{\lambda_n} \right)^2 \lambda_n = \infty$. In other words even though $R(\Lambda_1) = R(\Lambda_2)$ when $\Lambda_1\Lambda_2 = \Lambda_2\Lambda_1$, the conditions given in Theorem 2 need not be fulfilled always.

We shall finally take up the discussion of the general case when p_1 and p_2 are equivalent Gaussian measures with means m_1 and m_2 and dispersion operators Λ_1 and Λ_2. Let $\Lambda = \frac{1}{2}(\Lambda_1 + \Lambda_2)$. Then $m_1 - m_2 = \delta \epsilon R(\Lambda^{\frac{1}{2}})$ and $\Lambda_1 = \Lambda_2^{\frac{1}{2}} S \Lambda_2^{\frac{1}{2}}$ for some bounded self-adjoint S with $S \geqslant k.I$ for a constant $k > 0$ and $\operatorname{tr}(S-I)^2 < \infty$. It is easy to prove that there are bounded self-adjoint operators S_1 and S_2 with $S_1 \geqslant k_1.I$, $S_2 \geqslant k_2.I$ and $\operatorname{tr}(S_1-I)^2 < \infty$ and $\operatorname{tr}(S_2-I)^2 < \infty$ satisfying $\Lambda = \Lambda_1^{\frac{1}{2}} S_1 \Lambda_1^{\frac{1}{2}}$ and $\Lambda = \Lambda_2^{\frac{1}{2}} S_2 \Lambda_2^{\frac{1}{2}}$. It follows from this that $R(\Lambda_1^{\frac{1}{2}}) = R(\Lambda_2^{\frac{1}{2}}) = R(\Lambda^{\frac{1}{2}})$. If we now write q_1 and q_2 for the Gaussian measures with the same mean $m = \frac{1}{2}(m_1 + m_2)$ and dispersion operators Λ_1 and Λ_2, then it is an easy consequence of the above remarks and Theorem 5.1 that p_1, p_2, q_1 and q_2 are all equivalent. We are now in a position to prove our final theorem.

Theorem 6.3 : *Let p_1 and p_2 be equivalent Gaussian measures with means m_1, m_2 and dispersion operators Λ_1 and Λ_2. Let $m = \frac{1}{2}(m_1 + m_2)$ and $\delta = m_1 - m_2$. Suppose that* (a) $R(\Lambda_1) = R(\Lambda_2)$ (b) $(\Lambda_1^{-1} - \Lambda_2^{-1})\Lambda_2^{\frac{1}{2}}$ *extends to a bounded operator M defined over all of X with $tr(M^*M) < \infty$* (c) $\delta \epsilon R(\Lambda_1)$. *Then the closure A of $\Lambda_1^{-1} - \Lambda_2^{-1}$ exists, is a closed densely defined symmetric operator such that $p_2(D(A)+m) = 1$, and*

$$L(x) = -\frac{1}{2}(A(x-m), (x-m)) + \frac{1}{2}(x, \Lambda_1^{-1}\delta + \Lambda_2^{-1}\delta)$$

$$-\frac{1}{2}(m, \Lambda_1^{-1}\delta + \Lambda_2^{-1}\delta) - \frac{1}{8}(\delta, A\delta) - \frac{1}{2}\log|S|$$

for almost all x.

Proof : Since p_1, p_2, q_1 and q_2 are all equivalent we have

$$L = \log (dp_1/dq_1)+\log (dq_1/dq_2)-\log \ (dp_2/dq_2)$$

p_1 and q_1 are Gaussian measures with means m_1 and m and dispersion operator Λ_1. Moreover $m_1-m = \frac{1}{2}\delta\epsilon R(\Lambda_1)$. Hence by Theorem 6.1

$$\log (dp_1/dq_1)(x) = \frac{1}{2}(x, \Lambda_1^{-1}\delta)-\frac{1}{2}(m, \Lambda_1^{-1}\delta)-\frac{1}{8}(\delta, \Lambda^{-1}\delta)$$

for almost all x. Similarly,

$$\log \ (dp_2/dq_2)(x) = -\frac{1}{2}(x, \Lambda_2^{-1}\delta)+\frac{1}{2}(m, \Lambda_2^{-1}\delta)-\frac{1}{8}(\delta, \Lambda_2^{-1}\delta)$$

for almost all x. Now since $dq_1/dq_2(x) = dq_1'/dq_2'(x-m)$ where q_1' and q_2' are Gaussian measures with means 0 and dispersion operators Λ_1 and Λ_2, Theorem 6.2 enables us to conclude that the closure A of $\Lambda_1^{-1}-\Lambda_2^{-1}$ exists, is a closed, symmetric densely defined operator with $q_2'(D(A)) = 1$. Consequently $q_2(D(A)+m) = 1$ and for almost all x

$$\log\left(\frac{dq_1}{dq_2}\right)(x) = -\frac{1}{2}(A(x-m), \ (x-m))-\frac{1}{2}\log \ |S|$$

Combining all the three expressions we obtain the required formula for L.

REFERENCES

CAMERON, R. H. and MARTIN, W. T. (1944) : Transformations of Wiener integrals under translations. *Ann. Math.*, **45**, 386-396.

——— (1945) : Transformations of Wiener integrals under a general class of transformation. *Trans. Amer. Math. Soc.*, **58**, 184-219.

FELDMAN, JACOB (1958) : Equivalence and perpendicularity of Gaussian processes. *Pacific Jour. Math.*, **8**, 699-708.

GELFAND, I. M. and YAGLOM, A. M. (1956) : Integration in functional spaces and its application in quantum physics. *Uspekhi. Matem. Nauk*, **11**, 77-114 (in Russian).

GRENANDER, U. (1952) : Stochastic processes and statistical inference. *Arkiv för Matematik* **1**, 195-277.

HAJEK, J. (1958) : On a property of the normal distribution of any stochastic process. *Czechoslovak Math. J.*, **8**, 610-618.

HELLINGER, E. (1909) : Neue bergrundung der theorie quadratisher formen von unendlichvielen veranderlichen. *J. fur. reine und angew. Mathematic*, 136, 210.

KRAFT, C. (1955) : Some conditions for consistency and uniform consistency of statistical procedures. *Univ. California Publ. Statist.* **2**, 125-142.

MAHALANOBIS, P. C. (1925) : Analysis of race mixture in Bengal. *Jour. Asiat. Soc.* (Bengal), **23**, 301.

——— (1937) : Normalisation of statistical variates and the use of rectangular co-ordinates in the theory of sampling distributions. Appendix, *Sankhyā*, **3**, 35-40.

PROHOROV, YU, V. (1956) : Convergence of random processes and limit theorems in probability. *Theor. Veroyat. I.ee. Prim.* **1**, 177-238 (in Russian).

RAO, C. R. (1954) : On the use and interpretation of distance function in statistics. *Bull. Int. Statist. Inst.*, **34**, Part II, 1-10.

RIESZ, F. and NAGY, B. Sz. (1952) : Lecons D'analyse fonctionnelle. Budapest.

SEGAL, I. E. (1958) : Distributions in Hilbert space and canonical systems of operators. *Trans. Amer. Math. Soc.*, **88**, 12-41.

SILOV, G. E. (1963) : Integration in infinite dimensional spaces and the Wiener integral. *Uspekhi Matem. Nauk.* 18, 99-120.

SNEATH, P. H. A. and ROBERT, H. SOKAL (1962) : Numerical Taxonomy. *Nature*, **193**, 855-860.

STONE, M. H. (1932) : Linear transformations in Hilbert space and their applications to analysis. *Amer. Math. Soc.*, Colloquium Publications.

LESS VULNERABLE CONFIDENCE AND SIGNIFICANCE PROCEDURES FOR LOCATION BASED ON A SINGLE SAMPLE : TRIMMING/WINSORIZATION 1

By JOHN W. TUKEY and DONALD H. McLAUGHLIN
Princeton University[1]

SUMMARY. The vulnerability of Student's t, insofar as efficiency and power are concerned, leads to consideration of substitutes. Among the most promising are ratios of trimmed means to square roots of suitable quadratic forms involving the same order statistics. Matching, across underlying distributions, of ratios of average of denominator to variance of numerator leads to selection of the Winsorized sum of squared deviations as the basis for a denominator. The resulting *trimmed t* should prove more useful when the amount of trimming is made to depend on the individual sample in a suitably prescribed manner. Exact critical values for the resulting *tailored t* seem to require Monte Carlo computation, but use of a simple modified denominator for trimmed t allows us to use the conventional t tables as a reasonable approximation.

1. Introduction

One of us (Tukey, 1962, p.16) has already summarized the advantages of the class of symmetric distributions as a natural first step in our progress from statistical techniques understood and known to be useful for Gaussian (= normal) distributions alone to techniques understood and useful in very much more general situations. Once we are firmly established with statistical techniques understood and known to be useful for symmetrical distributions, it will be time to take a further step. But the problem of adequate mastery of the symmetrical-distribution case is enough for the moment.

Indeed it seems enough for the present to confine ourselves to techniques which are not only symmetric in their action on the class of all samples, but are symmetric in their action on individual samples. Such a restriction is clearly both less important and more easily removed than the restriction to symmetrical distributions.

The present account confines itself further : (i) to the single-sample-for-location problem and (ii) to the first steps of a specific approach to that problem. It describes the results of certain easily accessible calculations, and outlines what appear to be the plausible next steps, as well as indicating a little of what is known, or believed to be true, beyond this scope.

2. The normal and the pathological : which is which ?

The Gaussian or Laplacian distribution, to the physicist the Maxwellian distribution, has long been known to the statistician as the normal distribution. However little noticed such a commonly-used name becomes, shades of its original meaning continue to cling—distributions that are not normal are, by at least slight implication, pathological. In the early stages of development or exploration of a particular aspect

[1] Research supported by U.S. Army Research Office (Durham). Reproduction in whole or part for any purpose of the U.S. Government permitted.

of statistics or data analysis, such an attitude may promote progress. But in a well-developed area such an attitude can only be respectable if it reflects the facts—if the *usual* is at least close to the "normal" in behaviour. This is not the case in almost all of the instances of data analysis which the writers—and, they believe, most practising statisticians, have come in contact. The typical distribution of errors and fluctuations has a shape whose tails are longer than that of a Gaussian distribution. (See Tukey (1960) for more extended discussion, and Mandelbrot (1960, 1961, 1962) for newer instances arising in economics.)

It is the Gaussian distribution that has to be regarded as somewhat pathological from the standpoint of practice. And distributions with shorter tails, while they do occur, are rather more pathological. Thus frequency of occurrence directs our attention to longer-tailed distributions.

There is another, quite distinct and independent, reason for emphasizing long-tailed distributions. The prevalence of minimax-loss approaches to uncertainty is not an accident. We all tend to have more interest in avoiding a large loss than in obtaining a large gain. If procedures optimum for Gaussianity are used against long-tailed distributions they tend to behave poorly, both relatively and absolutely. Their quality is usually not only far from optimal (for the specific situation) but also of far lower quality than would have been the case if the underlying distribution were normal.

Against short-tailed distributions, on the other hand, procedures optimum for Gaussianity are not infrequently relatively poor but absolutely good, in the sense that, while the optimum procedure for the specific situation would do much better, the performance of the Gaussianly optimum procedure will be better for short-tailed distributions than for Gaussian ones.

The use of s^2 as an indicator of scale is, of course, an outstanding instance of the behaviour just discussed, both for short-tailed and long-tailed distributions.

Accordingly, it seems appropriate to begin by giving major attention to symmetric distributions with tails longer than the Gaussian.

3. Measures of quality and insensitivity

Statistical procedures are identified among the more general procedures of data analysis, which themselves may or may not be based on a probability model, by the fact that they take explicit account of uncertainty. From a narrow viewpoint, the most important aspect of such a procedure is its *validity*, the extent to which any associated statements of probability are correct, or at least conservative. Does the nominal significance level really apply? Does the formal confidence interval have (at least) the asserted probability of covering the true value? When such questions are asked about the behaviour for other underlying distributions of a procedure calibrated for Gaussianity, the conventional term is "robustness." If we wish to be clear and specific, we should—and shall—speak of "robustness of validity."

But only a slightly broader view causes us to ask of a procedure not only "Is it valid?" but "Is it efficient?". While controlling its rate of error, does it do as well, say as much, extract as much from the data, etc., as it can? And when we ask such questions about the behaviour for other underlying distribution of a procedure first developed for (near) Gaussianity we are asking about "robustness of efficiency."

So long as we are to continue to use Gaussian underlying distributions as the standard of calibration and the natural starting point, thus assuring validity in the Gaussian situation,—and this seems likely to be a long, long, time—these arguments suggest that we should give major attention for the present to

(1) efficiency for Gaussian distributions,

(2) robustness of validity for long-tailed symmetrical distributions,

(3) robustness of efficiency for long-tailed symmetrical distributions, where the first probably deserves by far the least attention of the three.

4. Location from a single sample : competitors and challenges

If we are given a single (random) sample, $y_1, y_2, \ldots, y_n$ of observations from $dF(y-\mu)$ where $F(v)+F(-v) = 1$ (so that the distribution of y is symmetric around μ), a statistical technique which permits tests of significance also provides confidence statements, and vice versa.

The most classical technique for this problem is one of many sorts of uses of Student's t. Its rather moderate robustness of validity has been studied by a number of workers (Pearson, 1929; Rietz, 1939; Gayen, 1949; Bradley, 1952). (In two-sample and simple analysis-of-variance situations Student's t is much more robust.) What we know about its behaviour can be summarized as follows :

(a1) The average value of the square of its denominator is in fixed ratio to the variance of its numerator, independent of the underlying distribution.

(a2) Its robustness of validity for symmetric underlying distributions is moderate, being quite high for significance or diffidence levels of 30-40% (Gayen, 1949) but not as satisfactory for the usual 5%, 1%, etc. levels. (Its behaviour for unsymmetric underlying distributions is much less satisfactory.)

(a3) Contrary to most naive intuitions, confidence and significance will be over-estimated by Student's t, not when the underlying distribution is longer-tailed (as Gayen, 1949; Bradley, 1952; and Wonnacott, 1963 all agree) but rather when the underlying distribution is shorter-tailed (as Rider, 1929; Perlo, 1933; Laderman, 1939; and Gayen, 1949 all agree).

(a4) Its robustness of efficiency is subject to serious question, especially since a single wild-appearing observation can seriously affect both $\bar{y}$ and s.

(a5) The method of its calculation can easily be extended (or analogized) to a very wide variety of situations without requiring changes in (Gaussian-theory) critical values for this reason.

(a6) It provides confidence limits with little more effort than significance tests.

(a7) If the underlying distribution should be Gaussian, these procedures will be optimal according to almost every criterion.

Toward the other extreme we find techniques which *can* be based on ordering deviations (of y's from a contemplated central value M) according to magnitude and basing the test (or confidence interval) upon the pattern of signs of deviations, which we will call the *sign-configuration*. This is most frequently and simply done by sorting the ranks of one sign in some simple way.

Scoring each rank with its rank number is called the one-sample Wilcoxon or signed-rank procedure and was introduced by Wilcoxon (1945, 1946, 1947, 1949). (For more available expositions see Moses, 1953; or Siegel, 1956, where the procedure is applied to differences of paired observations.) In his thesis, Walsh (1947, 1949) demonstrated a result equivalent to the fact that the probability of obtaining any configuration of signs of deviations (when the deviations are ranked by magnitude) is the same for all symmetric underlying distributions. (For a clarification of the relationship of his results to Wilcoxon's see Walsh, 1959.)

The most important aspects of our knowledge of such sign-configuration procedures can be summarized as follows :

(b1) Their robustness of validity is perfect for symmetric distributions.

(b2) Their robustness of efficiency has not been adequately studied.

(b3) The method of their calculation does not seem to be trivially extendable to more general situations (such as regression coefficients); new tables of critical values seem almost certain to be needed in any such extension.

(b4) The calculation of confidence intervals requires appreciably more effort than significance testing, although trial-and-error is not required (Moses, 1953), a simple graphical approach sufficing. (In more general situations, trial-and-error seems likely to be necessary.)

(b5) If the underlying distribution should be Gaussian, the efficiency of various of these procedures will be very high (Klotz, 1962), the loss of efficiency in comparison with the "optimum" t-procedure being almost negligible in this situation.

If we were only concerned with the single-sample problem, and were able to settle the question of robustness of efficiency for some sign-configuration procedure favourably, it would be reasonable to argue that such a procedure was a reasonable choice for routine work. Its only major defect would be its tendency, because of difference in labour, to cause its users to stop with significance tests in many instances where it would be profitable for them to push on to confidence intervals.

If we are to seek a new procedure, it should be one combining many of the relative advantages of t-procedures and sign-configuration procedures. As an ideal,

possibly utopian, we might seek a procedure with these properties for symmetric underlying distributions :

(c1) Its robustness of validity is high.

(c2) Its robustness of efficiency is satisfactorily large.

(c3) Its method of calculation can be rather easily extended to a wide variety of situations with little change in critical values.

(c4) It provides confidence intervals almost as easily as significance tests.

(c5) If the underlying distribution should be Gaussian, its efficiency is high.

If we could find a procedure which met, or came close to meeting, these specifications, it would be an outstanding candidate for adoption as the method of choice for routine use.

5. A DESIRABLE DIRECTION OF EXPLORATION

The criteria toward which we hope to make progress are diverse in kind and character—it would be unrealistic to expect any formal optimization procedure to actually lead us toward our goal. Accordingly, we have a choice between trying to modify the things we understand, or seeking to be struck by the lightning of a purely new approach. While waiting for lightning, we may as well proceed with modification, which appears to be more easily accomplished starting with Student's t.

When reached through a formal optimization procedure, Student's t arises as a single, integrated creation—and thus offers little guidance for modification. But Student's t did not first arise in such a way. Its numerator and denominator have very different conceptual origins, namely :

(d1) The numerator came first, as an intuitively effective point estimate of deviation from contemplated value.

(d2) The denominator followed, as something which reflected (indeed, was an estimate of) the variability of the previously chosen numerator.

One road—to the writers the currently most promising road—toward the goal set out in (c1) to (c5) is thus to begin by choosing a modified numerator, and then to seek a matching denominator.

In the case of Student's t, where attention was concentrated on an underlying Gaussian distribution — as was wholly appropriate when breaking new ground— "matching" needed only to refer to this situation. The fact that the denominator continued to estimate the variance of the numerator for all distributions was only a bonus, albeit one that proved to be very important. In the present approach, "matching" must refer to at least a modestly wide variety of symmetric distributions.

Granted that both frequency of occurrence and intensity of danger should cause us to give particular attention to longer-tailed distributions, our modification of the numerator must lie in the direction of attaching less weight to extreme—more precisely extreme-appearing—observations.

6. Trimming and Winsorizing

There are two sorts of simple modifications of an arithmetic mean which especially deserve consideration in this context, both for reasons of simplicity and for reasons arising from analyses of mathematical models to be reported elsewhere.

Given n *ordered* observations

$$y_1 \leqslant y_2 \leqslant \dots \leqslant y_n \qquad \dots \quad (1)$$

the (unweighted or equally weighted) arithmetic mean $\bar{y}$ or $y_{\bullet}$ is given by

$$y_{\bullet} = \frac{1}{n}\,(y_1+y_2+\dots+y_n) = \Sigma\, y_j/\Sigma 1. \qquad \dots \quad (2)$$

If $n = g+h+g$ (this mode of expression is chosen instead of $n = 2g+h$ to stress ordering), the g-times (symmetrically) trimmed mean y_{Tg} is given by

$$y_{Tg} = \frac{1}{n-2g}\,(y_{g+1}+y_{g+2}+\dots+y_{n-g}) = \sum_{(Tg)} y_j/\sum_{(Tg)} 1 \qquad \dots \quad (3)$$

and is the arithmetic mean of the set of h numbers obtained by dropping both the g lowest and the g highest values from the y_j. Clearly y_{Tg} pays less attention to extreme values than does $y_{\bullet}$, which may be regarded as the 0-fold trimmed (= untrimmed) mean.

A less intuitive contender is the g-times (symmetrically) Winsorized mean, y_{Wg}, given by

$$y_{Wg} = \frac{1}{n}\,(g{\cdot}y_{g+1}+y_{g+1}+y_{g+2}+\dots+y_{n-g}+g{\cdot}y_{n-g}) = \sum_{(Wg)} y \,/ \sum_{(Wg)} 1 \qquad \dots \quad (4)$$

which is the arithmetic mean of the n values obtained by replacing (i) each of the g lowest y's by the value of the nearest other y, namely y_{g+1} and (ii) each of the g highest y's by the value of the nearest other y, namely y_{n-g}. Again we have paid less attention to individual extreme y's, but we have managed not to divert our attention from the "tails" of the sample so thoroughly. Instead of replacing each deleted y_j by y_{Tg}, which is one reasonable interpretation of the calculation of y_{Tg}, since

$$y_{Tg} = \frac{1}{n}\,(g{\cdot}y_{Tg}+y_{g+1}+y_{g+2}+\dots+y_{n-g}+g{\cdot}y_{Tg}), \qquad \dots \quad (5)$$

we have only replaced each deleted y by the nearest retained y. (As noted by Dixon (1960), this procedure has been called Winsorization in honour of Charles P. Winsor, who actively sponsored its use in actual data analysis.)

We shall find it convenient to continue to use the notation just illustrated throughout the discussion that follows, namely

(e1) Replacement of a subscript by a "$\bullet$" indicates a simple arithmetic mean.

(e2) Replacement of a subscript by "Tg" indicates a g-times trimmed mean.

(e3) Replacement of a subscript by "Wg" indicates a g-fold Winsorized mean.

(e4) The indication "(Tg)" on a summation sign indicates (g-times) trimmed summation = summation over the subscripts remaining after the sample of y's is trimmed g times on each tail.

(e5) The indication "(Wg)" on a summation indicates (g-times) Winsorized summation = summation over the subscripts for which the y's were not deleted, repeating each extreme undeleted subscript $g+1$ times and each other subscript once.

If we were concerned with but one amount of trimming or one amount of Winsorization we could well make use of simpler notations. (One such has been suggested at the end of (b2) on page 12 of Tukey, 1960.) But when concerned with several values of g it seems advisable to use a more explicit notation.

7. Choice of numerator

For underlying distributions whose shapes are very close to Gaussian, the Winsorized means are less variable than trimmed means. While the efficiency for Gaussianity of trimmed means is quite high, the fractional loss being crudely $2g/3n$ (corresponding to efficiency of about 2/3 for the median), that of the corresponding Winsorized means is much higher. At the other extreme, where very long-tailed distributions are involved, trimmed means are clearly more efficient than Winsorized means. Where does the transition take place ?

At the time when a previous account was written, preliminary analysis suggested a moderately broad scope for the Winsorized mean (cf. Tukey, 1962, p. 18). Now that further analysis (to be reported elsewhere) has been carried out, it would appear that the trimmed mean is likely to be more widely useful than had been supposed. As we shall see, this change in interpretation makes the present programme more attractive.

8. Criteria of matching

In striving to "match" denominators to a given numerator, we must choose a criterion of matching. The natural, and we believe reasonable choice, is to begin by following the example of Student's t and ask that

average value of denominator squared

and variance of numerator

should be in constant proportion over as broad a spectrum of symmetrical distributions as is reasonably convenient. This is, again, *only a first step*. Once we find a denominator which matches well in this sense, we are ready to calculate critical values of the ratio for various symmetrical distributions, and learn whether its validity is really robust.

The numerators we consider are linear combinations of order statistics; their variances will be linear combinations of variances and covariances of order statistics. And the squared denominators are likely to be quadratic functions of order statistics; their average values will depend on averages, variances, and covariances of order statistics. Accordingly we naturally begin by turning to those symmetrical distributions for which low moments of order statistics are available.

Three distributions are outstanding in this regard :

(f1) the rectangular distribution, for which low moments are available in closed form,

(f2) the Gaussian distribution, for which 1st and 2nd moments are available for sample sizes $\leqslant 20$ (Teichroew, 1956; Sarhan and Greenberg, 1956; see alternatively Teichroew, 1962),

(f3) one long-tailed distribution—the lambda distribution with $\lambda = -0.1$ —called the "special distribution" by Hastings, Mosteller, Tukey, and Winsor (1947), for which they provided 1st and 2nd moments for sample sizes $\leqslant 10$.

In addition to these distributions, published tables of low moments of order statistics from symmetric distributions seem restricted to

(f4) the isosceles triangular distribution (Sarhan, 1954, p. 320) for sample sizes up to 5, and

(f5) the double exponential distribution (Sarhan, 1954, p. 320) for sample sizes up to 5.

In view of the central importance of the Gaussian distribution, and the ease of handling order statistic moments for the rectangular distribution, the natural course is to begin by trying to "match" for these two distributions and, once reasonable success is obtained, to check the match for the other distributions, giving special emphasis to the match for distributions longer-tailed than the Gaussian.

Work on the preparation of extensive tables of order statistic moments for lambda distributions is in progress. When these are available, somewhat better checks on matching will be possible. However, since the ultimate check is in terms of % points of the ratio rather than in terms of comparing individual values of numerator and denominator, the need for such further checks is not great.

9. Denominators matched to the trimmed mean : early trials

The denominator most naively associated with the trimmed mean, y_{Tg}, is (some multiple of) the formal standard deviation of the trimmed sample, whose square is proportional to the (g-times) trimmed sum of squared deviations (SSD)

$$\mathrm{SSD}_{Tg} = \sum_{(Tg)} (y_i - y_{Tg})^2. \qquad \ldots \quad (6)$$

The most convenient of the ratios that must be nearly constant if we are to have matching is

$$\frac{\text{ave } \mathrm{SSD}_{Tg}}{\text{var } y_{Tg}} = \text{divisor}_1(g+h+g) \qquad \ldots \quad (7)$$

where the name "divisor" is justified by the fact that

$$\sqrt{\frac{\mathrm{SSD}_{Tg}}{\text{divisor}_1\,(g+h+g)}} \qquad \ldots \quad (8)$$

is the natural normalization to an actual denominator for use with y_{Tg}. When we investigate the behaviour of

$$\frac{\text{Gaussian divisor}_1\,(g+h+g)}{\text{rectangular divisor}_1\,(g+h+g)} \qquad \cdots \quad (9)$$

where "Gaussian" and "rectangular" specify the distributions for which "ave" and "var" are calculated = the underlying distributions for which the appropriate forms of (8) yield denominators the average of whose square equals the variance of y_{Tg}, we find that (9) is moderately, but probably not satisfactorily close to unity, as Table 1 shows. For more detail, see McLaughlin and Tukey (1961).

TABLE 1. VALUES OF THE RATIO (9) FOR SELECTED g AND h

h=size of sample after trimming	g=number of observations trimmed from each end					
	1	2	3	5	8	9
2	1.009	1.007	1.005	1.003	1.002	1.001
3	1.002	1.002	1.002	1.001	1.001	
4	.995	.996	.997	.998	.999	
5	.987	.989	.991	.995		
9	.966	.967	.971	.979		
10	.963	.962	.967	.976		
15	.951	.947				
18	.948					

Reflection upon these results showed that, for better matching, the denominator should be modified in such a way as to give more attention to the outlying portions of the sample. This led to trials of various alternatives such as using SSD_{Tg^*}, where $g^* < g$, with y_{Tg} so that a less-trimmed sample provided the denominator. Trial and consideration eventually led to investigation of

$$\mathrm{SSD}_{Wg} = \sum_{(Wg)} (y_i - y_{Wg})^2 \qquad \cdots \quad (10)$$

the equally-many-times *Winsorized* sum of squares of deviations as a basis for a denominator to be used with the *trimmed* mean y_{Tg}.

The matching of the corresponding divisor for different distributions

$$\text{divisor}_2\,(g+h+g) = \frac{\text{ave } \mathrm{SSD}_{Wg}}{\text{var } y_{Tg}} \qquad \cdots \quad (11)$$

has to be examined. The first check is again of

$$\frac{\text{Gaussian divisor}_2\,(g+h+g)}{\text{rectangular divisor}_2\,(g+h+g)} \qquad \cdots \quad (12)$$

with the results shown in Table 2.

TABLE 2. VALUES OF THE RATIO (12) FOR SELECTED g AND h

h=size of sample after trimming	g=number of observations trimmed from each end					
	1	2	3	5	8	9
2	1.0092	1.0071	1.0053	1.0031	1.0017	1.0014
3	1.0052	1.0045	1.0036	1.0023	1.0013	
4	1.0037	1.0033	1.0027	1.0018	1.0011	
5	1.0031	1.0023	1.0021	1.0015		
9	1.0026	1.0021	1.0015	1.0009		
10	1.0026	1.0020	1.0015	1.0009		
15	1.0024	1.0020				
18	1.0023					

The line for $h = 2$, where only the 2 central observations are retained in either the trimmed or Winsorized samples, must be the same in Tables 1 and 2. Elsewhere in Table 2, only one entry rises as much as 0.5% above unity. (As compared with a fall of more than 5% for the t-denominator.) And since divisor$_2(g+h+g)$ will appear under a square root, this corresponds to a suggested difference in critical value between rectangular and Gaussian of 1 part in 400. (As much as almost 1 in 200 when only the two central values survive trimming.) For the direction of easy computation but of lesser importance, the direction of shorter tails than Gaussian, the suggested behaviour of divisor$_2$ $(g+h+g)$ is close to excellent. What of the other side?

Table 3 presents the available values for

$$\frac{\text{special divisor}_2\,(g+h+g)}{\text{rectangular divisor}_2\,(g+h+g)} \quad \ldots \quad (13)$$

where "special" refers to the lambda distribution with $\lambda = -0.1$, (cf. Hastings, Mosteller Tukey, and Winsor (1947)), which can be roughly thought of as a t distribution with 5 degrees of freedom. Most of the values are close to 1.01, a value which suggests a 0.5% difference in critical values between the rectangular and the special.

TABLE 3. VALUES OF (13) FOR AVAILABLE g AND h

h=size of sample after trimming	g=number of observations trimmed from each end			
	1	2	3	4
2	1.025	1.018	1.013	1.010
3	1.016	1.012	1.009	
4	1.013	1.009	1.007	
5	1.011	1.008		
6	1.010	1.007		
7	1.010			
8	1.010			

The remaining easy comparisons are made in Table 4, which offers no reason to change the conclusions already reached.

TABLE 4. OTHER RATIOS OF VALUES OF divisor$_2$ $(1+h+1)$

distribution compared with the rectangular	h=size of sample after trimming	distribution divisor$_2$ / rectangular divisor$_2$
Isoceles triangular	2	1.007
	3	1.014
double exponential	2	1.041
	3	1.037

The "suggestion" of a somewhat larger divisor for longer-tailed distributions requires some consideration at this point. What we are finding is that the divisor, defined as

$$\frac{\text{ave (denominator)}^2}{\text{var numerator}} \qquad \dots \quad (14)$$

increases slightly as we pass from a very short-tailed distribution to a quite long-tailed one. In the case of Student's t, this ratio does not change at all, but Gayen's results suggest that, for modest numbers of degrees of freedom, the corresponding critical values change (decrease) by several times the fractions with which we are concerned. Since we would not expect this effect to be so prominent for trimmed and Winsorized statistics, it seems likely that the matching behaviour of SSD_{Wg}, as a denominator, at least so far as divisor$_2$ $(g+h+g)$ is concerned, is all that we can ask at this stage.

11. Approximation to the denominator—trimmed t

Depending upon the numerical behaviour of the divisor, we have a number of choices in putting the resulting procedure into approximate practice. If its behaviour is simple enough, we may calculate a "trimmed t" using a convenient approximation to the divisor, and then compare the results with, as an approximation, the (normal-theory) critical values of Student's t, or, more precisely, with modified critical values appropriate to the precise distribution (say on normal theory) of "trimmed t". Whether or not the behaviour of the divisor is simple, we can always choose a convenient working divisor and then tabulate the appropriate critical values. Clearly the first of these possibilities is the more desirable. Does divisor$_2$ behave simply?

Table 5 shows the ratio of the normal-theory values of divisor$_2$ to $h(h-1)$, all values falling between 1.00 and 1.02. (For rectangular-theory values, see Section 19.) Clearly we will do quite well to use $h(h-1)$ as the working divisor, especially when we recall that using a slightly undersized divisor corresponds to using slightly longer-confidence intervals, and is thus slightly conservative.

TABLE 5. RATIO OF NORMAL THEORY divisor_2 TO $h(h-1)$

h=size of sample after trimming	g=number of observations trimmed from each end					
	1	2	3	5	8	9
2	1.009	1.007	1.005	1.003	1.002	1.001
3	1.016	1.015	1.013	1.009	1.006	
4	1.016	1.016	1.015	1.011	1.008	
5	1.015	1.016	1.015	1.012		
9	1.010	1.012	1.012	1.011		
10	1.010	1.011	1.011	1.010		
15	1.007	1.008				
18	1.006					

The next question, of course, has to do with the approximate distribution of the result, particular in the vicinity of the conventional tail areas. To help us with this problem there are two pieces of information :

(g1) Dixon and Tukey (1963) have studied the approximate distribution of Winsorized t, where y_{Wg} (rather than y_{Tg}) is combined with s_{Wg}. The results of this study indicate a quite Student's-t-like distribution, with the best fit obtainable for a number of degrees of freedom typically somewhat less than $h-1$(= the number corresponding to a sample of size equal to the trimmed sample).

(g2) R. A. Jensen has made some preliminary experimental sampling and Monte Carlo investigations into the critical values of t_T. These suggest, rather than indicate, that these critical values may be rather close to those of Student's t for $h-1$ degrees of freedom.

Accordingly, if we wished to use trimmed t on an approximate basis (but see Sections 14 to 16), we would calculate

$$t_{Tg} = \frac{y_{Tg}-M}{\sqrt{\text{SSD}_{Wg}/h(h-1)}} \qquad \ldots \quad (15)$$

where M is a contemplated value for the center of the distribution sampled, and refer the result to Student's t on $h-1$ degrees of freedom.

And we could, as a next step, proceed to a more precise determination of actual critical values for (15).

12. Matching to the Winsorized mean: a query

Turning now to the Winsorized mean, analogy with the results just described leads us to begin with denominators which pay more attention to the tails of the sample than does the numerator. Working with the values of the trimmed sample alone, an attempt to match drives us rapidly to the (g-times) inner range.

$$W_{Tg} = y_{n-g} - y_{g+1} = W_{Wg} \qquad \ldots \quad (16)$$

The corresponding divisor

$$\text{divisor}_3\,(g+h+g) = \frac{\text{ave}(W_{Tg}^2)}{\text{var}\; y_{Wg}} \qquad \ldots \quad (17)$$

does not appear to be so satisfactorily constant as we change the underlying distribution, as Table 6 shows.

TABLE 6. VALUES OF $\frac{\text{Gaussian divisor}_3\,(g+h+g)}{\text{rectangular divisor}_3\,(g+h+g)}$

h=size of sample after trimming	1	2	3	5	8	9
2	1.009	1.007	1.005	1.003	1.002	1.001
3	.962	.985	.989	.994	.997	
4	.961	.963	.971	.982	.990	
5	.950	.945	.953	.968		
9	.961	.914	.900	.922		
10	.970	.914	.903	.913		
15	1.022					
18	1.066					

The most natural ways to seek improvement are (i) the use of observations *not in* y_{Wg} to help assess its variability, (ii) the use of a denominator in which differences among central observations are *subtracted* from W_{Tg} in order to further emphasize the behaviour of the ends of the trimmed sample. Both of these have strong heuristic disadvantages : the first because we may lose the advantage of getting wholly free of the observations excluded from the trimmed sample, with the consequence that, in very long-tailed distributions, we may improve our numerator rather more than we improve our estimate of its performance; the second because such subtractions, if effective, must tend to decrease the relative stability of the denominator (measured, if you will, in "effective degrees of freedom").

Whether either of these approaches, or some other, such as using

$$W_{Tg} + A \cdot |y_{Wg} - y_{Tg}| \qquad \ldots \quad (18)$$

for a suitable value of A, will prove satisfactory, seems likely to be better investigated by working with actual distributions of ratios and the corresponding critical values rather than with ratios of moments. If this be so, it will be well to gain experience first with the distributional behaviour of t_{Tg}.

13. POWER OF TRIMMED t

In connection with a study of Monte Carlo methods adapted to statistical problems of sampling from non-normal distributions, Wonnacott (1963) has investigated some selected aspects of the power of trimmed t both when the underlying distribution is Gaussian and when it is a symmetrical Johnson (1949) distribution with low moments corresponding to a t-distribution with 6 or 4.7 degrees of freedom. In general his results are as would have been expected, though the comparison of Student's t and Wilcoxon-Walsh procedures is, surprisingly, somewhat unfavourable to the latter.

A few highlights of Wonnacott's (1963) comparison of powers for non-normal underlying distributions are these :

(h1) Singly trimmed t for $n = 10$ is more powerful than Student's t for the Johnson distribution with moments matching t_6.

(h2) Five-times trimmed t for $n = 20$ is very close to Student's t for the same underlying distribution.

(h3) Three-times trimmed t for $n = 10$ is almost as powerful as the nearest Wilcoxon-Walsh procedure for the Johnson distribution with moments matching $t_{4\cdot 7}$.

In general, the prospects for the effectiveness of t_{Tg} seem very good.

14. WHAT SHOULD WE EXPECT TO BE USED ?

Once we have obtained all desired critical values of t_{Tg}, what are we to do in practice? Let us suppose the error rate (here = significance or diffidence level) at which we are going to work has been fixed, once for all. This assumption is of course unrealistic, but it enables us to look more clearly at the issues which concern us most immediately. We may as well also fix $n = g+h+g$.

Let, then,

$$a_g = \text{critical value of } t_{Tg} \qquad \dots \quad (19)$$

and consider a man with a single sample of y's. He has a wide variety of choices. He may take any of :

$$\begin{array}{ll} \textit{point estimate} & \textit{interval estimate} \\ y_{\bullet} & y_{\bullet} \pm a_0\sqrt{\text{SSD}/n(n-1)} \\ y_{T1} & y_{T1} \pm a_1\sqrt{\text{SSD}_{W1}/(n-2)(n-3)} \\ y_{T2} & y_{T2} \pm a_2\sqrt{\text{SSD}_{W2}/(n-4)(n-5)} \\ y_{T3} & y_{T3} \pm a_3\sqrt{\text{SSD}_{W3}/(n-6)n-7)} \end{array} \qquad \dots \quad (20)$$

We know that, if his underlying (symmetrical) distribution behaves "reasonably" (in quotes because we do not yet know what is "reasonable" and what is not), systematic use *without exception* of any *single* one of the interval estimates will offer him validity, those with larger g probably being more robust in this regard. But our motivation in entering upon this whole question was to improve efficiency while maintaining validity. And we are almost certain that the qualitative behaviour of relative efficiency will appear as in Figure 1.

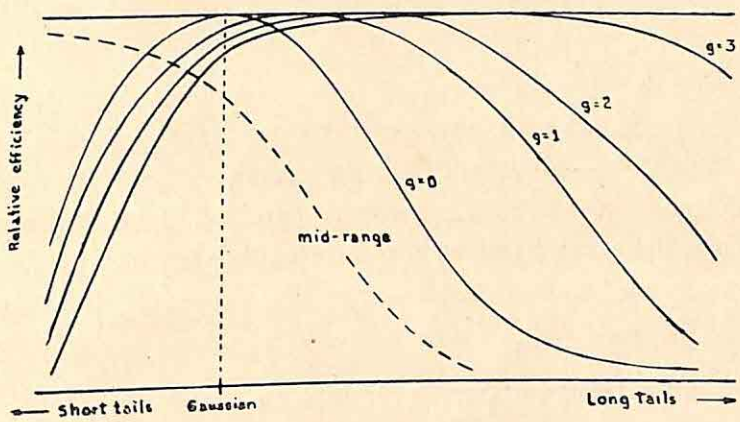

Fig. 1. Anticipated qualitative behaviour of relative efficiency.

Knowing this, it is most unlikely that the user will be content to use any single value of g *without exception*. When his samples are long-tailed, thus appearing to come from long-tailed distributions, he will want to use values of $g > 0$. When they look "Gaussian" he will wish to use $g = 0$ (=Student's t). (And it is conceivable that some will feel that they sometimes really do have an underlying distribution with shorter tails, and will wish to try to take advantage of this by occasional use of a numerator more of the character of the mid-range. But this, for the reasons discussed in Section 5, is well treated as a second-order perturbation.)

15. Is this wise ?

At first glance it looks as though such a user has started upon a dangerous course. All those who have handled much data know, often largely instinctively, how dangerous it would be to trust a single small sample to tell us with any precision about the shape of the distribution from which it came. And it tends to appear that this is what a man is doing when he picks a g in the light of the specific sample to which he is to apply the chosen procedure.

Let us turn to a simpler, classical situation. What of the man with a sample of, say, size 3 from a distribution known to be Gaussian? He has an estimate of σ^2 based on 2 degrees of freedom. If he believes that he "knows" σ^2 from other evidence, and acts accordingly, he is often in bad trouble. If he makes a significance test or states a confidence interval as if σ^2 were known, he will indeed be incurring much

larger risks than he claims. It has been 55 years since Student (1908) showed us the way out of this dilemma. We have only to plan to use an estimate, s^2, of σ^2 based on these observations, to admit that this choice will be fallible, and to ask how must we readjust the critical values to allow for this fallibility. Doing this for the use of s^2 for σ^2, which takes us to the distribution of Student's t, is a familiar operation, now regarded as logically simple. The only possible difficulties are computational ones associated with the calculation of appropriate critical values, once the rule for choosing the estimate of σ^2 is fixed. (Thus, for example, the answers for Lord's t (Lord, 1947, 1950) where this is based upon range, are somewhat different from those for Student's t.)

The situation for allowance for long tails is a similar one. Once we fix a way for the user to choose a given value of g—and decide what the *form* of the adjustment to the critical value will be—no logical difficulty remains. There is a system of critical values which will enable a man who behaves as specified, *without exception*, to make statements which will be valid (for appropriate underlying distributions). The only difficulty is a purely computational one—find the *amount* of the needed change in critical values.

16. Individually trimmed, t = tailored t

The selection of the exact procedure for choosing a value of g, so long as this procedure is sensible, is a matter of little importance. At first glance there are at least two plausible alternatives, namely :

$$\text{Choose that } g \text{ which minimizes } \mathrm{SSD}_{W_g}/h(h-1) \qquad \ldots \quad (21)$$

and

$$\text{Choose that } g \text{ which minimizes } a_g \cdot \sqrt{\mathrm{SSD}_{W_g}/h(h-1)}. \qquad \ldots \quad (22)$$

However, it is only for distributions with the very longest tails that it will be sensible to choose values of g that discard almost all the sample values. (The cost of discarding so many sample values will not be paid in terms of the stability of y_{T_g}, which cannot be made worse than the stability of the median—which is excellent in sampling from long-tailed distributions—but rather in terms of the stability of the denominator—stability of the estimate of variability of y_{T_g}.) The simplest means of avoiding difficulty in this connection will be to introduce a fixed function $G(n)$ and choose g according to :

$$\text{Choose that } g \leqslant G(n) \text{ which minimizes } \mathrm{SSD}_{W_g}/h(h-1) \qquad \ldots \quad (23)$$

or

$$\text{Choose that } g \leqslant G(n) \text{ which minimizes } a_g \cdot \sqrt{\mathrm{SSD}_{W_g}/h(h-1)}. \qquad \ldots \quad (24)$$

The choice between (23) and (24), *for a given* $G(n)$, is not likely to have a substantial effect on the adjustment required for the critical values. Nor is it likely to have any substantial effect upon the way in which relative efficiency depends on length of tail. The choice between (23) and (24) will almost surely be based upon considerations pertinent to the user.

There are three such which seem likely to be of major importance:

(j1) It is slightly easier to use (23), since the square-rootings and multiplications by a_g required in the criterion of (24) are avoided.

(j2) The use of (24) may be more palatable, since one begins by seeking out the g for which the naive limits of (20) are closest together, and one retains this best-seeming value of g.

(j3) The use of (24) makes it very easy to make a conservative correction to the confidence limits set by a man who has used the naive limits of (20), perhaps without revealing how he chose g.

The strength of (j3) is of course greatest, when the change in critical values takes the form

$$a_g \to b_{n,G} \cdot a_g \qquad \ldots \quad (25)$$

so that the resulting confidence limits are

$$y_{Tg} \pm b_{n,G} \cdot a_g \cdot \sqrt{\mathrm{SSD}_{Wg}/h(h-1)} \qquad \ldots \quad (26)$$

where g is chosen by (24).

17. Purely a matter of selection ?

One attitude toward the problems we have just been discussing would be that they are purely matters of allowance for selection of procedure, another routine instance of a broad problem which routinely faces us in the analysis of data : There is a single set of data, and several alternative procedure by which it could be analyzed. While it is possible to retain validity by pledging that one will always use a single procedure, it is clear that failure to do something to adapt the procedure to the data leads to a loss of effectiveness. But if we naively select the procedure that appears to work best in each specific instance, we are exposing ourselves to a certain loss of validity.

There are almost always two ways out of such a bind, the one we propose to adopt, namely calculating adjusted critical values which allow for the choice, by a prescribed rule, of the apparently most appropriate procedure, and one which is rather in the spirit of Robbins' empirical Bayes techniques (Robbins, 1951, 1955; see also Neyman, 1962). In the second approach, one would regard an individual batch of data as one of a family of such batches, and plan to borrow information from the other batches to determine how we are to proceed with the first. (This general procedure is of course adopted daily, often at a very general level, by every working statistician whose familiarity with data of a certain class or classes help him to choose among procedures for its analysis.)

If we wish to appear "whiter than driven snow" insofar as selection is concerned, we may decide to choose the procedure to be used upon a given batch of data solely upon the evidence offered by the *other* batches of the family. If (i) the batches are independent, and (ii) their assignment to a "family" was without regard to their behaviour, even the most vehement and inquisitive seeker for selection bias could not ascribe any to the way in which the procedure was selected.

Why did we not propose such an approach to the selection of g? Mainly, we believe, because of a feeling about what will eventually be discovered to be the case. Specifically, suppose that the shape of the underlying distribution is fixed and that we are applying some procedure, or mixture of procedures, to samples drawn from distributions with this fixed shape. On the basis of available evidence and insights it is reasonable to suppose that one can do better, even after making due allowance for the $b_{n,G}$ factor that will then be required, by using different values of g for samples of different apparent long-tailedness, than by using any one fixed value of g.

For a specific shape of distribution, it is a factual question whether "mixing the g's" is advantageous or not. In time we should know the answer for certain specific distributions.. In the meantime, we can do no better than follow our best judgement.

One could, if one wished, repeat an analogous argument about $G(n)$, which, in contrast to g, we have implicitly proposed to determine from a whole large family of batches of data. We have chosen to avoid making such an argument, and feel quite happy, on our present knowledge and insight, in making a sharp distinction between :

(k1) $G(n)$ to be picked on the evidence of a *family* of batches, and

(k2) g to be picked, subject to $g \leqslant G(n)$, on the evidence of *the single* batch in question.

To what extent this distinction is made because of a deeper understanding of the role of g, and to what extent it reflects a real distinction between what the user can gain from individualized choice of g in comparison with the individualized choice of $G(n)$ is hard to say.

18. Required next steps

What then are the next steps to be taken in making tailored t properly available for use? Some of them, surely, are these :

(l1) Determine the values of a_g for an adequate net of values of g and n, and suitable error rates.

(l2) Choose a reasonable function $G(n)$ and determine the values of b_{Gn} for a suitable pattern of values of g, n, and error rate. (A considerably sparser pattern may suffice for adequate interpolation.)

(l3) Investigate the power of the resulting procedure, both for a Gaussian underlying distribution and for longer-tailed underlying distributions. Comparison with both Student's t and sign-configuration procedures will be in order.

(l4) Make a start on extending the ideas and insights gained in the single-sample situation to more general situations.

Once reasonable progress has been made with (l1) and (l2), tailored t will be fully useable. We would expect to recommend it for routine use at that time.

So far as one can see, all the steps just mentioned demand Monte Carlo techniques for their solution, although we would not exclude the possibility of an analytic attack on some of the simpler ones. It should be emphasized that by Monte Carlo we do not mean naive experimental sampling, where one merely draws samples from the postulated underlying distribution and calculates a trimmed-t value from each, thus building up an empirical distribution approaching that of trimmed t like $n^{-\frac{1}{2}}$. Such naive procedures waste computational effort, and drastically reduce the accuracy of results that can be reached with plausible amounts of effort. Instead one must plan to use as much as possible of one's knowledge and insight in designing a modified sampling scheme whose results estimate a number known to be the same as the number estimated by the naive procedure. (See Kahn, 1956 for general discussion, and Arnold, Bucher, Trotter, and Tukey, 1956, or Wonnacott, 1963 for specific examples.)

While we plan to work on these problems at Princeton, we would welcome activity by others.

19. Algebra for the rectangular case

In closing we should set down the algebra which shows that

$$\text{rectangular divisor}_2(g+h+g) \doteq h(h-1). \qquad \dots \quad (27)$$

For a rectangular distribution with ends at 0 and 1 we have the well-known results for the low moments of the order statistics $y_1 \leqslant y_2 \leqslant y_3 \leqslant \dots \leqslant y_n$

$$\left.\begin{aligned} \text{ave } y_i &= \frac{i}{n+1} \text{ for all } i \\ \text{cov}(y_i, y_j) &= \frac{i(n+1-j)}{(n+1)^2(n+2)} \text{ for } i<j \end{aligned}\right\} \qquad \dots \quad (28)$$

whence

$$\text{var}(y_j-y_i) = \frac{(j-i)(n+1)-(j-i)^2}{(n+1)^2(n+2)} \qquad \dots \quad (29)$$

and

$$\text{ave}(y_j-y_i)^2 = \frac{(j-i+1)(j-i)}{(n+1)(n+2)} \qquad \dots \quad (30)$$

which depends upon $j-i$ and n alone as would be expected from the symmetric distribution of equivalent blocks.

In view of the general identity

$$2{\cdot}\Sigma 1 \cdot \Sigma(u_i-u_.)^2 = \Sigma\Sigma(u_j-u_i)^2 \qquad \dots \quad (31)$$

we have, using the Winsorized observations as the u_i

$$\begin{aligned} 2n \cdot \text{SSD}_{W_g} = {} & 0+2g \sum_{g+1}^{n-g}(y_j-y_{g+1})^2 + \sum_{g+1}^{n+g}\ \sum_{g+1}^{n-g}(y_j-y_i)^2 \\ & +2g\Sigma(y_{n-g}-y_i)^2+2g^2(y_{n-g}-y_{g+1})^2+0, \end{aligned} \qquad \dots \quad (32)$$

whence, setting $i = g+r$, $j = g+s$, and using (30)

$$2n(n+1)(n+2) \cdot \text{ave SSD}_{W_g} = 4g \sum_1^h (r-1)(r) + \sum\sum_1^h (r-s)(r-s+1) + 2g^2 \cdot (h-1)h \ \dots \quad (33)$$

Now $\Sigma\, r(r-1) = (h+1)(h)(h-1)/3$. And the double sum can be easily evaluated by considering a sample of size h, from the same rectangular distribution, and the corresponding untrimmed SSD, say SSD(z), for which

$$2h \cdot \text{SSD}(z) = \Sigma\Sigma(z_r-z_j)^2 \qquad \ldots \quad (34)$$

$$\text{ave SSD}(z) = (h-1)\sigma^2 = \frac{h-1}{12} \qquad \ldots \quad (35)$$

Now write (33) for $n = h$ and $g = 0$, finding

$$2\cdot h(h+1)(h+2) \text{ ave SSD}(z) = \sum_1^h\Sigma(r-s)(r-s+1) \qquad \ldots \quad (36)$$

but the left-hand side equals

$$2h(h+1)(h+2)\,\frac{(h-1)}{12} = \frac{1}{6}\,(h-1)\,h(h+1)(h+2) \qquad \ldots \quad (37)$$

so that

$$\frac{1}{6}\,(h-1)\,h(h+1)(h+2) = 2(h)(h+1)(h+2) \text{ ave SSD}(z) = \Sigma\Sigma(r-s)(r-s+1). \quad \ldots \quad (38)$$

$$2n(n+1)(n+2) \text{ ave SSD}_{W_g} = 4g\,\frac{(h-1)\,h(h+1)}{3} + \frac{1}{6}\,(h-1)\,h(h+1)(h+2)$$
$$+2g^2\,(h-1)h \qquad \ldots \quad (39)$$

and
$$\text{ave SSD}_{W_g} = \frac{(h-1)h}{12n(n+1)(n+2)}\,[(h+1)(h+2)+8g(h+1)+12g^2]$$

$$= \frac{(h-1)h}{12n(n+1)(n+2)}\,[3n^2-(h-2)(2n+1)] \qquad \ldots \quad (40)$$

We turn now to
$$h^2 \text{ var } y_{T_g} = \sum_{(T_g)}\Sigma \text{ cov } (y_i, y_j) \qquad \ldots \quad (41)$$

and notice that, if $r < s$ $\quad (n+1)^2(n+2) \text{ cov } (y_{g+r},\, y_{g+s})$

$$= (g+r)(n+1-g-s) = (g+r)(g+h+1-s)$$
$$= g^2+g(r-s)+g(h+1)+(h+1)^2(h+2) \text{ cov } (z_r, z_s)$$
$$= g^2-g\cdot|r-s|+g(h+1)+(h+1)^2(h+2) \text{ cov } (z_r, z_s) \qquad \ldots \quad (42)$$

where the requirement that $r < s$ can be lifted for the last form, so that, using (41)

$$h^2\,(n+1)^2(n+2) \text{ var } y_{T_g} = h^2\cdot g^2 - g\sum_1^r\Sigma\,|r-s| + h^2(h+1)g+(h+1)^2(h+2)(h^2 \text{ var } z_\bullet) \quad (43)$$

and, since $\quad h^2\cdot \text{var } z_\bullet = h/12$ and $\Sigma\Sigma\,|r-s| = h(h^2-1)/3$

we have
$$\text{var } y_{T_g} = \frac{(h+1)^2(h+2)+12h(h+1)g-4(h^2-1)g+12hg^2}{12h(n+1)^2(n+2)}$$

$$= \frac{h(3n-2h+3)+2}{12h(n+1)(n+2)}. \qquad \ldots \quad (44)$$

Now

$$\frac{\text{ave SSD}_{W_g}}{h(h-1)\,\text{var}\, y_{Tg}} = \frac{h(3n^2-2nh+4n-h+2)}{n(h(3n-2h+3)+2)}$$

$$= 1+\frac{(n-h)\,(h-2)}{n(3(n-h)h+(h+1)(h+2))} \quad \cdots \quad (45)$$

which will usually be close to 1.

References

Arnold, H. J., Bucher, B. D., Trotter, H. F. and Tukey, J. W. (1956): Monte Carlo techniques in a complex problem about normal samples. *Symposium on Monte Carlo Methods.* (Edited by H. A. Meyer), John Wiley and Sons, Paper 9, 80–88.

Bradley, R. A. (1952): The distribution of the t and F statistics for a class of non-normal populations. *Virginia J. Science,* **3**, 1–32.

Dixon, W. J. (1960): Simplified estimation from censored normal samples. *Ann. Math. Stat.,* **31**, 385–391.

Dixon, W. J. and Tukey, J. W. (1963): Approximate behaviour of the distribution of Winsorized t (Trimming/Winsorization 2). In manuscript.

Gayen, A. K. (1949): The distribution of 'Student's' t in random samples of any size drawn from non-normal universes. *Biometrika,* **36**, 353–369.

Hastings, C. Jr., Mosteller, F., Tukey, J. W. and Winsor, C. P. (1947): Low moments for small samples: a comparative study of order statistics. *Ann. Math. Stat.,* **18**, 413–426.

Johnson, N. L. (1949): Systems of frequency curves generated by methods of translation. *Biometrika,* **36**, 149–176.

Kahn, H. (1956): Use of different Monte Carlo sampling procedures. *Symposium on Monte Carlo Methods.* (Edited by H. A. Meyer), John Wiley and Sons, Paper 15, 146–190.

Klotz, J. (1962): Small sample power and efficiency for the one sample Wilcoxon and normal scores tests. Paper read at meeting of the Institute of Mathematical Statistics, Minneapolis, 7 Sept. 1962, 17 pp.

Laderman, J. (1939): The distribution of "Student's" ratio for samples of two items drawn from non-normal universes. *Ann. Math. Stat.,* **10**, 376–379.

Lord, E. (1947): The use of range in place of standard deviation in the t-test. *Biometrika,* **34**, 41-67. (Correction, **39**, 442).

——— (1950): Power of the modified t-test (u-test) based upon range. *Biometrika,* **37**, 64–77.

Mandelbrot, B. (1960): The Pareto-Levy law and the distribution of income. *Int. Econ. Rev.,* **1**, 79–106.

——— (1961): Stable Paretian random functions and the multiplicative variation of income. *Econometrica,* **29**, 517–543.

——— (1962): The variation of certain speculative prices. *IBM Research Note NC*-**87**, 111.

McLaughlin, D. H. and Tukey, J. W. (1961): The variance of symmetrically trimmed samples from normal populations, and its estimation from such trimmed samples. (Trimming/Winsorization 1). *Stat. Tech. Res. Group,* (Princeton University) *Tech. Rept.,* **42**, 24.

Moses, L. (1953): Non-parametric methods. H. M. Walker and J. Lev's *Statistical Inference,* Henry Holt, Chapter 18, 426–450.

Neyman, J. (1962): Two breakthroughs in the theory of statistical decision making. *Rev. Internat. Statist. Inst.,* **30**, 11–27.

Pearson, E. S. assisted by Adyanthaya, N. K. (1929): The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. *Biometrika,* **21**, 251–289.

PERLO, V. (1953) : On the distribution of Student's ratio for samples of three drawn from a rectangular distribution. *Biometrika*, **25**, 203–204.

RIDER, P. R. (1929) : On the distribution of the ratio of mean to standard deviation in small samples from non-normal universes. *Biometrika*, **21**, 124–143.

RIETZ, H. L. (1939) : On the distribution of the "Student" ratio for small samples from certain non-normal populations. *Ann. Math. Stat.*, **10**, 265–274.

ROBBINS, H. (1951) : Asymptotically subminimax solutions of compound statistical decision problems. *Proc. 2nd Berkeley Symp. Math. Stat. Prob.*, 131–148.

——— (1955) : An empirical Bayes' approach to statistics. *Proc. 3rd Berkeley Symp. Math. Stat. and Prob.*, **1**, 157–164.

SARHAN, A. E. (1954) : Estimation of the mean and standard deviation by order statistics. *Ann. Math. Stat.*, **25**, 317–328.

SARHAN, A. E. and GREENBERG, B. G. (1956) : Estimation of location and scale parameters by order statistics from singly and doubly censored samples. Part I, the normal distribution up to samples of size 10. *Ann. Math. Stat.*, **27**, 427–451.

SIEGEL, S. (1956) : *Non-Parametric Statistics for the Behavioral Sciences*, McGraw-Hill, 75–83.

STUDENT (1908). On the probable error of a mean. *Biometrika*, **6**, 1–25. Reprinted in "*Student's*" *Collected Papers*, Biometrika Office, Paper 2, 11–34.

TEICHROEW, D. (1956) : Tables of expected values of order statistics for samples of size 20 and less from the normal distribution. *Ann. Math. Stat.*, **27**, 410–426.

——— (1962) : Tables of lower moments of order statistics for samples from the normal distribution. *Contributions to Order Statistics* (Edited by A. E. Sarhan and B. G. Greenberg), Chapter 10B, 190–205, John Wiley.

TUKEY, J. W. (1960) : A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics: Essays in honour of Harold Hotelling* (Edited by I. Olkin *et al.*), 448–485.

——— (1962) : The future of data analysis. *Ann. Math. Stat.*, **33**, 1–67 and 812.

WALSH, J. E. (1947) : Some significance tests for the median which are valid under very general conditions. Ph.D. Thesis, Princeton University.

——— (1949) : Some significance tests for the median which are valid under very general conditions. *Ann. Math. Stat.*, **20**, 64–81.

——— (1959) : Comments on "The simplest signed-rank tests". *J. Amer. Stat. Ass.*, **54**, 213–224.

WILCOXON, F. (1945) : Individual comparisons of grouped data by ranking methods. *Biometrics Bulletin* (later *Biometrics*) **1**, 80–83.

——— (1946) : Individual comparisons of grouped data by ranking methods. *J. Econ. Entomology*, **39**, 269–270.

——— (1947) : Probability tables for individual comparisons by ranking methods. *Biometrics*, **3**, 119–122.

——— (1949 and later editions) : *Some Rapid-Approximate Statistical Methods*, American Cyanamid Company, Stamford Research Laboratories.

WONNACOTT, T. H. (1963) : A Monte-Carlo method of obtaining the power of certain tests of location when sampling from non-normal distributions. Ph.D. Thesis, Princeton University.

Paper received : February, 1963.

BIOMETRIKA

Volume 50, Parts 3 and 4 **December, 1963**

CONTENTS

The subscription, *payable in advance*, is 54*s* (or $8.00), per volume (including postage). Cheques should be made payable to *Biometrika*, crossed "*A/c Biometrika Trust*" and sent to the Secretary, Biometrika Office, Department of Statistics, University College, London, W. C. 1. All foreign cheques must be drawn on a Bank having a London agency.

Issued by THE BIOMETRIKA OFFICE, University College, London.

JOURNAL OF THE

ROYAL STATISTICAL SOCIETY

Series B (Methodological)

40*s* each Part. Annual Subscription £3.2*s* (US $9). Post free

CONTENTS OF VOL. 25, PART 2, 1963

Obtainable through any bookseller or direct from the

ROYAL STATISTICAL SOCIETY

21 Bentinck Street, London, W. 1.

THE ANNALS OF MATHEMATICAL STATISTICS

(Founded by H. C. Carver)

The Official Journal of the Institute of Mathematical Statistics

CONTENTS OF VOLUME 34, NO. 4, DECEMBER, 1963

Subscription Rate: $ 10.00 per year

Address orders for subscriptions and back numbers to the Secretary, Institute of Mathematical Statistics, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

ECONOMETRICA

Vol. 31, No. 4 **October, 1963**

CONTENTS

Subscription price for nonmembers: $ 12.00 per year in the United States and Canada; $ 10.00 per year in all other Countries. Single copies $ 3.00
Graduate Students recommended by a faculty member: $ 4.00 per year.

Subscriptions should be sent to

The Secretary
ECONOMETRIC SOCIETY
Box 1264, Yale Station, New Haven, Connecticut, U.S.A.

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Editors: P. C. MAHALANOBIS, C. R. RAO

SERIES A, VOL. 25 DECEMBER 1963 PART 4

THE BEHRENS-FISHER TEST WHEN THE RANGE OF THE UNKNOWN VARIANCE RATIO IS RESTRICTED*

By WILLIAM G. COCHRAN

Harvard University

SUMMARY. In some applications of the Behrens-Fisher test, it is reasonable to suppose that the unknown variance ratio σ_1^2/σ_2^2 must exceed a known quantity λ_1/λ_2. The effect of this restriction on the significance levels of the test statistic d is examined by computing the probability that d exceeds the tabulated significance level, in the restricted region, for 6, 12, and 24 degrees of freedom in s_1^2 and s_2^2 and probability levels of 5% and 1%. If the F test made from the data gives strong support to the supposition that $\sigma_1^2/\sigma_2^2 > \lambda_1/\lambda_2$, the disturbance to the Behrens-Fisher significance levels is minor, but otherwise it can be substantial and can lie in either direction. The practical use of these results is discussed.

1. INTRODUCTION

In the Behrens-Fisher problem we are given a comparison x, normally distributed with mean μ and variance $(\sigma_1^2+\sigma_2^2)$. We also have independent estimates s_1^2 of σ_1^2 and s_2^2 of σ_2^2, based on n_1 and n_2 degrees of freedom. The problem is to test the null hypothesis that μ has some stated value, usually zero. The test criterion is $d = (x-\mu)/\sqrt{s_1^2+s_2^2}$.

Although the Behrens-Fisher test is intended to involve no assumptions about the variance ratio σ_1^2/σ_2^2, there are practical problems in which it is reasonable to assume that σ_1^2/σ_2^2 must exceed some known value. To cite an example discussed by Fisher (1941), the mean of a large sample of (n_1+1) crude measurements of some physical quantity may be compared with the mean of a smaller number (n_2+1) of refined measurements, in order to examine whether the two processes are biased relative to one another. If σ_a, σ_b are the standard deviations of the populations of crude and refined measurements, respectively, the assumption $\sigma_a/\sigma_b > 1$ appears justified. In applying the Behrens-Fisher test to this problem, x is the difference between the sample means and

$$s_1^2 = \frac{s_a^2}{(n_1+1)} \quad : \quad s_2^2 = \frac{s_b^2}{(n_2+1)}.$$

Consequently, the restriction $\sigma_a^2/\sigma_b^2 \geqslant 1$ implies that $\sigma_1^2/\sigma_2^2 > (n_2+1)/(n_1+1)$.

* This research was supported by the Contract Nonr 1866(37) with the Office of Naval Research, U. S. Navy Department. The computations were done at the Harvard Computing Center under grant NSF-GP-683 by the National Science Foundation.

This situation occurs also in certain comparisons in the analysis of split-plot or nested experiments. Factor A, with a levels, is applied to relatively large plots or experimental units in some standard design. Each unit is divided into b equal subunits, to which are applied the b levels of a second factor B. The mathematical model used in the analysis postulates that the error e_{ijk} on the subunit receiving the i-th level of A, the j-th level of B and lying in the k-th replication has mean zero and variance σ^2. The errors e_{ijk}, $e_{i'j'k'}$ on subunits in different units are assumed independent, but errors e_{ijk}, $e_{ij'k}$ on two subunits in the same unit have a correlation ρ. It follows that the error variance of a unit total, when divided so as to express it on a subunit basis, is

$$\sigma_a^2 = \sigma^2\{1+(b-1)\rho\}. \qquad \dots \quad (1)$$

However, the variance of the difference between two subunits in the same unit, also on a single subunit basis, is

$$\sigma_b^2 = \sigma^2(1-\rho). \qquad \dots \quad (2)$$

An unbiased estimate s_a^2 of σ_a^2, obtained from the analysis of variance of unit totals, is used for testing the main effects of A. The subunit analysis supplies an independent estimate s_b^2 of σ_b^2 for tests involving main effects of B and AB interactions.

If interactions are present and B is a qualitative factor, the experimenter may wish to test comparisons like $(a_2b_1)-(a_1b_1)$; that is, the difference between the means for two A levels at the same level of B. In terms of the model, the error variance of such comparisons is $2\sigma^2/r$, where r is the number of replications. From equations (1) and (2) an unbiased estimate of this variance is

$$\frac{2\{s_a^2+(b-1)s_b^2\}}{rb}.$$

If x is the mean difference between (a_2b_1) and (a_1b_1), the Behrens-Fisher test may be used if we take

$$s_1^2 = \frac{2s_a^2}{rb} \qquad s_2^2 = \frac{2(b-1)s_b^2}{rb} \qquad \dots \quad (3)$$

giving to s_1^2 the d.f. in s_a^2 and to s_2^2 those in s_b^2.

In many split-plot or nested experiments there is reason to believe that $\rho > 0$. From equations (1) and (2) this implies that $\sigma_a^2/\sigma_b^2 > 1$ and hence in (3) that $\sigma_1^2/\sigma_2^2 > 1/(b-1)$.

The objective of this paper is to make a preliminary exploration of the disturbance produced in the Behrens-Fisher test when the range of σ_1^2/σ_2^2 is restricted in this way.

2. Numerical example

The following notation covers the type of problem illustrated by the two preceding examples. Let σ_a^2 and σ_b^2 be the variances in the two populations in which we are prepared to assume that $\sigma_a^2/\sigma_b^2 > 1$. Let $f' = s_a^2/s_b^2$ be the corresponding estimated variance ratio based on n_1 and n_2 degrees of freedom. The value of f' will be known from the data. Further, let

$$\sigma_1^2 = \lambda_1\sigma_a^2 \qquad \sigma_2^2 = \lambda_2\sigma_b^2 \qquad \dots \quad (4)$$

where σ_1^2 and σ_2^2 are the variances that enter into the Behrens-Fisher test, λ_1 and λ_2 being known numbers that depend on the problem in question. Then the restriction may be written

$$\phi = \frac{\sigma_1^2}{\sigma_2^2} > \frac{\lambda_1}{\lambda_2}. \qquad \dots \quad (5)$$

For later calculations it is convenient to rewrite this as

$$\frac{\phi}{f} > \frac{1}{f'} \qquad \dots \quad (6)$$

where $f = s_1^2/s_2^2 = \lambda_1 f'/\lambda_2$.

The effect of this additional information about $\phi = \sigma_1^2/\sigma_2^2$ on the significance levels of d may be examined from a result due to Fisher, who showed that the unrestricted significance levels of d can be computed in the following way. The probability that d exceeds a specified value is found first on the assumption that $f = s_1^2/s_2^2$ and $\phi = \sigma_1^2/\sigma_2^2$ are both known. The average value of this probability over all possible values of ϕ from 0 to ∞ is then calculated by assigning to ϕ/f its fiducial distribution for known f, this distribution being the tabular F distribution with n_2 and n_1 d.f.

To obtain the frequency distribution of d when f and ϕ are both given, Fisher notes that $x/\sqrt{\sigma_1^2+\sigma_2^2}$ follows the standard normal distribution and that

$$\frac{n_1 s_1^2}{\sigma_1^2} + \frac{n_2 s_2^2}{\sigma_2^2}$$

follows that χ^2 distribution with (n_1+n_2) d.f. Hence a variate that follows Student's distribution with (n_1+n_2) d.f. is

$$\begin{aligned} t &= \frac{x\sqrt{n_1+n_2}}{\sqrt{\left(\frac{n_1 s_1^2}{\sigma_1^2}+\frac{n_2 s_2^2}{\sigma_2^2}\right)(\sigma_1^2+\sigma_2^2)}} \\ &= \frac{d\sqrt{s_1^2+s_2^2}\,\sqrt{n_1+n_2}}{\sqrt{\left(\frac{n_1 s_1^2}{\sigma_1^2}+\frac{n_2 s_2^2}{\sigma_2^2}\right)(\sigma_1^2+\sigma_2^2)}} \\ &= \frac{d\sqrt{1+f}\,\sqrt{n_1+n_2}}{\sqrt{\left(n_2+\frac{n_1 f}{\phi}\right)(1+\phi)}}. \end{aligned} \qquad \dots \quad (7)$$

By means of this relation, the probability that d exceeds any specified value, for given f and ϕ, is read from the Student t-table with (n_1+n_2) degrees of freedom. This probability is then averaged over the fiducial distribution of ϕ/f from 0 to ∞.

When it is known that $\phi/f > 1/f'$, the natural modification of the Behrens-Fisher technique is to average this probability only over the values of ϕ/f that exceed $1/f'$. This is the method that will be adopted in this paper.

From the preceding discussion, writing $u = \phi/f$, the probability that d exceeds the value d_α when we confine ourselves to the restricted region $u > 1/f'$ is given by the expression

$$\int_{1/f'}^{\infty} \frac{u^{\frac{n_2}{2}-1} P\{|t| > d_\alpha g(u)\} du}{(n_1+n_2u)^{\frac{n_1+n_2}{2}}} \Bigg/ \int_{1/f'}^{\infty} \frac{u^{\frac{n_2}{2}-1} du}{(n_1+n_2u)^{\frac{n_1+n_2}{2}}} \quad \dots \quad (8)$$

where

$$g(u) = \left\{ \frac{u(n_1+n_2)(1+f)}{(n_1+n_2u)(1+uf)} \right\}^{\frac{1}{2}}$$

and $P\{|t| > d_\alpha g(u)\}$ is the two-tailed probability that Student's t with (n_1+n_2) degrees of freedom exceeds $d_\alpha g(u)$.

As an illustration, consider a split-plot experiment with $n_1 = 6$, $n_2 = 12$, $b = 2$. These parameters hold if the main units are arranged in a 4×4 latin square, each unit having two subunits. In the solid line in Figure 1 the probability that d exceeds 2.301 (the 5% Behrens-Fisher value for $n_1 = 6$, $n_2 = 12$, $f = 1$) is plotted against the percentiles of the distribution of ϕ/f. The average probability (area under the line) is of course 0.05. Throughout most of the range of ϕ/f the probability lies below 0.05, this being required to compensate for very high and low values of ϕ that make the probability rise steeply towards 1 at both ends. Note that probabilities above 0.05 are mostly contributed by *high* values of ϕ/f. This happens whenever $n_2 > n_1$.

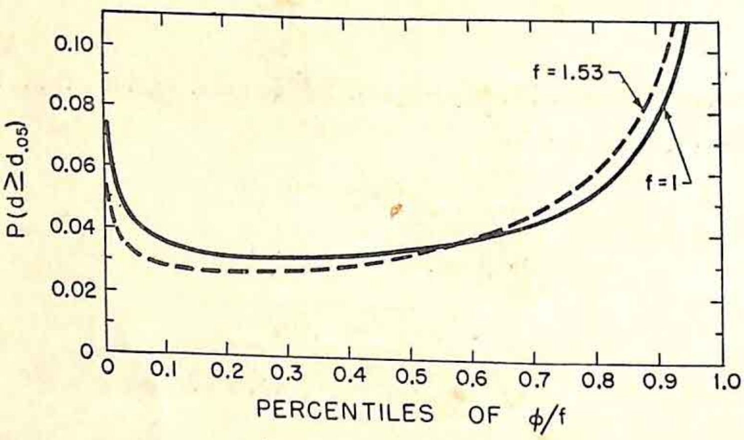

Fig. 1. Probability of exceeding $d_{.05}$ for given f and ϕ/f ($n_1 = 6$, $n_2 = 12$).

Suppose that in the split-plot experiment the investigator finds $s_a^2 = s_b^2$, i.e. $f' = 1$. For a split-plot, $f = f'/(b-1)$, so that since $b = 2$, $f = f' = 1$. Hence the restricted region over which the probability must be averaged is $\phi/f > 1$. This is approximately the region to the right of the median 0.5 on the abscissa. Eye inspection suggests that the average probability in the restricted region will exceed 0.05. Numerical integration gives 0.063.

The dotted line in Figure 1 shows the corresponding probabilities for $f = f' = 1.53$. Since this is the 25% significance level of f' it gives slightly greater support

to the idea that $\sigma_a^2/\sigma_b^2 > 1$ than does our previous choice of $f' = 1$. For $f' = 1.53$ the restricted region extends from 0.25 to 1 on the abscissa. The average probability in this region is 0.057. As f' increases further, the restricted region increases in size and the average probability moves toward 0.05. For f' lower than 1 the average probability is higher than that for $f' = 1$, being, for example, about 0.084 when $f' = 1/3$. (In each case the probabilities are those of exceeding the 5% value of d for $n_1 = 6$, $n_2 = 12$ and the appropriate f.) Similar results hold at the 1% level. The average probability in the curtailed region is 0.015 when $f' = 1$ and 0.012 when $f' = 1.53$.

3. Range covered by the computations

In order to investigate these effects more systematically, a series of calculations were made of the actual probability with which d exceeds $d_{.05}$ and $d_{.01}$, the tabulated Behrens-Fisher significance levels, in the restricted region. The values chosen for n_1 and n_2 were 6, 12 and 24, giving nine combinations. The quantity f' was taken at its 75%, 50%, 25% and 5% levels. With $f' = s_a^2/s_b^2$ at its 5% level, the data are confirming the investigator's idea that $\sigma_a^2/\sigma_b^2 > 1$, while when f' is at its 75% level, the data are tending to disagree with this apriori assumption. It was not thought necessary to investigate the situations in which f' is at its 1% or 0.1% levels, although such cases might be expected to occur commonly in practice when $\sigma_a^2/\sigma_b^2 > 1$. When f' is at these levels the restricted range is very close to the whole sample space of ϕ/f, so that the ordinary significance levels of d are unlikely to be much in error.

Expression (8) shows that the probability depends on f as well as on n_1, n_2 and f'. For the two-sample comparison in Section 1,

$$f = (n_2+1)f'/(n_1+1).$$

Thus f/f' is usually close to n_2/n_1, lying between this value and unity. In a split-plot experiment f/f' is easily seen to be n_1/n_2 if the main units are arranged in a completely randomized design. It is $an_1/(a-1)n_2$ when main units are in randomized blocks and $an_1/(a-2)n_2$ when main units form a latin square, where a is the number of levels of factor A. In all three split-plot cases f/f' lies between n_1/n_2 and unity.

For such applications it looks as if values of f/f' lying between n_1/n_2 and n_2/n_1 are primarily of interest. In more complex situations, however, the ratio f/f' need not bear any simple relation to n_1/n_2. Consequently the probabilities were computed for $f/f' = 0$, 1/4, 1/2, 1, 2 and ∞.

For any specific n_1, n_2, f and f' the ordinary Behrens-Fisher significance level d_α was first computed on the IBM 7090 by interpolation in the Fisher-Yates tables. Expression (8) was then obtained by numerical integration, using the trigonometric expansion of the integral of the t-distribution. I am greatly indebted to Michael Feuer who programmed and conducted the calculations. After debugging, computation of the 432 probability values took 17.9 minutes of machine time.

TABLE 1. PROBABILITY THAT d EXCEEDS THE BEHRENS-FISHER 5% LEVEL IN THE RESTRICTED REGION $\sigma_1^2/\sigma 1 > 1/f'$

f'	$f \to 0$	$f = f'/4$	$f = f'/2$	$f = f'$	$f = 2f'$	$f \to \infty$
			$n_1 = 6,\ n_2 = 6$			
75%	.012	.023	.034	.050	.070	.122
50%	.017	.029	.039	.050	.062	.082
25%	.026	.037	.043	.050	.055	.063
5%	.040	.046	.048	.050	.051	.052
			$n_1 = 6,\ n_2 = 12$			
75%	.024	.042	.055	.073	.092	.132
50%	.030	.044	.053	.064	.073	.087
25%	.036	.046	.052	.057	.060	.064
5%	.045	.049	.050	.051	.052	.052
			$n_1 = 6,\ n_2 = 24$			
75%	.035	.055	.069	.087	.105	.139
50%	.039	.053	.062	.071	.079	.090
25%	.043	.052	.056	.060	.063	.065
5%	.048	.050	.051	.052	.052	.053
			$n_1 = 12,\ n_2 = 6$			
75%	.007	.013	.019	.030	.045	.091
50%	.013	.021	.027	.036	.047	.070
25%	.023	.030	.036	.042	.048	.058
5%	.039	.044	.046	.048	.050	.052
			$n_1 = 12,\ n_2 = 12$			
75%	.019	.029	.037	.050	.065	.100
50%	.025	.035	.041	.050	.059	.075
25%	.033	.041	.045	.050	.054	.060
5%	.044	.047	.049	.050	.051	.052
			$n_1 = 12,\ n_2 = 24$			
75%	.030	.042	.051	.064	.078	.107
50%	.035	.045	.051	.059	.066	.078
25%	.041	.047	.051	.054	.058	.062
5%	.047	.049	.050	.051	.052	.052
			$n_1 = 24,\ n_2 = 6$			
75%	.004	.008	.012	.020	.032	.072
50%	.010	.016	.021	.029	.038	.061
25%	.020	.027	.032	.038	.044	.055
5%	.038	.042	.044	.047	.048	.051
			$n_1 = 24,\ n_2 = 12$			
75%	.014	.021	.027	.037	.048	.078
50%	.022	.029	.034	.041	.049	.065
25%	.031	.037	.041	.045	.050	.057
5%	.043	.046	.047	.049	.050	.051
			$n_1 = 24,\ n_2 = 24$			
75%	.026	.034	.041	.050	.060	.084
50%	.032	.039	.044	.050	.056	.068
25%	.039	.043	.047	.050	.053	.058
5%	.046	.048	.049	.050	.051	.052

TABLE 2. PROBABILITY THAT d EXCEEDS THE BEHRENS-FISHER 1% LEVEL IN THE RESTRICTED REGION $\sigma_1^2/\sigma_2^2 > 1/f'$

f'	$f \to 0$	$f=f'/4$	$f=f'/2$	$f=f'$	$f=2f'$	$f \to \infty$
			$n_1=6,\ n_2=6$			
75%	.0007	.0024	.0048	.0096	.0165	.0322
50%	.0014	.0037	.0062	.0100	.0137	.0186
25%	.0026	.0055	.0078	.0101	.0117	.0131
5%	.0059	.0084	.0094	.0100	.0103	.0105
			$n_1=6,\ n_2=12$			
75%	.0030	.0073	.0118	.0182	.0251	.0348
50%	.0040	.0082	.0112	.0147	.0172	.0194
25%	.0055	.0090	.0107	.0121	.0129	.0133
5%	.0080	.0097	.0102	.0104	.0105	.0105
			$n_1=6,\ n_2=24$			
75%	.0055	.0118	.0170	.0235	.0294	.0363
50%	.0065	.0112	.0142	.0168	.0185	.0197
25%	.0076	.0107	.0120	.0128	.0131	.0133
5%	.0091	.0102	.0104	.0105	.0105	.0105
			$n_1=12,\ n_2=6$			
75%	.0002	.0007	.0014	.0036	.0078	.0234
50%	.0006	.0016	.0028	.0053	.0087	.0160
25%	.0018	.0033	.0050	.0072	.0094	.0123
5%	.0053	.0072	.0083	.0092	.0099	.0104
			$n_1=12,\ n_2=12$			
75%	.0018	.0037	.0059	.0098	.0150	.0263
50%	.0028	.0051	.0071	.0100	.0129	.0172
25%	.0045	.0067	.0083	.0100	.0113	.0127
5%	.0075	.0089	.0095	.0100	.0103	.0105
			$n_1=12,\ n_2=24$			
75%	.0042	.0075	.0105	.0150	.0198	.0286
50%	.0054	.0083	.0104	.0129	.0151	.0180
25%	.0069	.0090	.0102	.0114	.0122	.0130
5%	.0088	.0097	.0101	.0103	.0105	.0105
			$n_1=24,\ n_2=6$			
75%	.0001	.0002	.0006	.0015	.0039	.0172
50%	.0003	.0008	.0015	.0031	.0058	.0135
25%	.0012	.0023	.0035	.0054	.0077	.0115
5%	.0048	.0064	.0075	.0086	.0094	.0103
			$n_1=24,\ n_2=12$			
75%	.0010	.0021	.0033	.0057	.0093	.0194
50%	.0020	.0035	.0049	.0071	.0096	.0146
25%	.0038	.0054	.0067	.0083	.0098	.0119
5%	.0072	.0083	.0089	.0096	.0100	.0103
			$n_1=24,\ n_2=24$			
75%	.0031	.0051	.0070	.0099	.0135	.0214
50%	.0045	.0064	.0080	.0100	.0120	.0155
25%	.0062	.0078	.0088	.0100	.0110	.0123
5%	.0085	.0093	.0097	.0100	.0102	.0104

4. Results

Tables 1 (5% level) and 2 (1% level) show the probability that d exceeds the 5% or 1% level in the Behrens-Fisher table when we make the additional assumption that $\sigma_a^2/\sigma_b^2 > 1$. The results are not too simple to summarize and digest, but the following points emerge.

(1) If one tries to guess the direction of the results intuitively, the easiest case is that in which $f = f'$, so that the restriction becomes $\sigma_1^2/\sigma_2^2 > 1$. It seems natural (at least to me) to guess that this restriction produces the same effect as that of an increase in s_1^2/s_2^2 on the ordinary Behrens-Fisher levels, because the additional information suggests that the stability of d will now depend more on the accuracy with which σ_1^2 is estimated than it does in the unrestricted case. When $n_1 < n_2$, an increase in s_1^2/s_2^2 raises the value of d required for 5% significance in the Behrens-Fisher tables for the range of values of n_1, n_2 considered here. Consequently the restriction should produce probabilities greater than 0.05 or 0.01 in Tables 1 and 2 when $f = f'$ and $n_1 < n_2$. Similarly, the restricted probabilities should be less than 0.05 or 0.01 when $f = f'$ and $n_1 > n_2$. These anticipations are verified in every case in Tables 1 and 2.

When $f = f'$ and $n_1 = n_2$, the same intuition suggests that the probabilities should not be disturbed. To the degree of accuracy shown in Tables 1 and 2 this happens in all 12 cases at the 5% level and in all but four cases at the 1% level. The discrepancies in these four cases are so small that they may be due to rounding errors in the calculations. I have tried to prove by integration that the probability for this set of cases remains exactly at 5% or 1% but have not succeeded, except for the case $f = f' = 1$ in which the result is obvious by symmetry.

(2) As anticipated, when f' is at the 5% level the probabilities remain close to those in the Behrens-Fisher tables except when f is very small or very large. As f' diminishes, the disturbance to the probabilities steadily increases, becoming very substantial when f' is at the 50% and 75% levels.

(3) For given f', n_1 and n_2 the probability increases monotonically with f, as can be verified mathematically.

(4) Although the panel of values of n_1 and n_2 is not large enough to suggest firm rules of interpolation against these values, it appears in all cases that for fixed n_1/n_2, the probability moves towards 0.05 or 0.01 as (n_1+n_2) increases.

Application to data

The primary reason for this investigation lies, of course, in the help that it might give to the investigator who wants to apply the Behrens-Fisher test under these restricted conditions. Unfortunately, the amount of help actually supplied is

limited. It is clear that for the sample sizes considered, the disturbance to the Behrens-Fisher probabilities can be substantial and can lie in either direction, and no simple rationalization of the whole pattern of results has occurred to me. A more extensive table of the 5% and 1% significance levels of d in the restricted region is perhaps called for, though as a four-variable table it would be inconvenient to use.

The result that the disturbance to the Behrens-Fisher values is minor (except for very small values of f relative to f') when f' is at or beyond the 5% level will often be all that the investigator needs to know. In looking for examples that might be typical of the split-plot case, I noticed that in the experiments reported in the well-known books by Snedecor, Federer, Bennett and Franklin, and Cochran and Cox, all the f' values were beyond the 0.5% level ($P < 0.005$) and three of the four were beyond the 0.1% level.

The result that there is either no disturbance or at most a trifling disturbance when $n_1 = n_2$ and $f = f'$ is also useful, since this applies to the comparison of two samples of equal sizes (and to a split-plot with two subunit treatments and main units completely randomized).

For the split-plot experiment with main units completely randomized, in randomized blocks, or in a latin square, we have, as noted previously, $n_1 \leqslant n_2$ and f/f' lying between n_1/n_2 and unity. In all such cases in Tables 1 and 2 with $n_1 < n_2$, the Behrens-Fisher significance level is too low for the restricted region, but usually by only a small amount. As an example, the experiment quoted in Goulden's book (1952) has $n_1 = 35$, $n_2 = 42$, $f = f' = 1.17$. This lies at about the 32% level. For $n_1 = 12$, $n_2 = 24$ in Table 1, the probability for $d_{.05}$ and $f = f'$ is 0.054 when f' is at the 25% level and 0.059 when f' is at the 50% level. For $n_1 = 21$, $n_2 = 42$ both probabilities presumably move towards 0.05, with a further move in this direction when $n_1 = 35$, $n_2 = 42$. Hence the probability in the restricted region may be expected to be only slightly above 0.05 in Goulden's example.

In the comparison of two samples of *unequal* sizes, $f/f' = (n_2+1)/(n_1+1)$, which will be approximately n_2/n_1 in most practical situations. Two cases may be distinguished. If $n_1 > n_2$, i.e. the less precise sample is larger, Tables 1 and 2 indicate that the probabilities are smaller than 0.05 or 0.01. For instance, in Table 1 with $n_1 = 24$, $n_2 = 6$, $f/f' = 1/4$, the 0.05 probabilities drop to 0.042, 0.027, 0.016 and 0.008. The disturbances can clearly be large if n_1 is much greater than n_2. The probabilities imply, of course, that a smaller value of d is needed for 5% significance than that given in the Behrens-Fisher tables. If $n_1 < n_2$, on the other hand, the probabilities are higher than the stipulated 0.05 and 0.01, as seen from $n_1 = 6$, $n_2 = 12$, $f/f' = 2$, and from $n_1 = 12$, $n_2 = 24$, $f/f' = 2$.

In conclusion, it is hoped that by using Tables 1 and 2 as illustrated above, the investigator can find out the direction of the disturbance to the Behrens-Fisher significance levels and obtain some idea as to whether the disturbance is likely to be minor or major.

REFERENCES

FISHER, R. A. (1939) : The comparison of samples with possibly unequal variances. *Ann. Eugen.*, **9**, 174-180.

——— (1941) : The asymptotic approach to Behrens' integral, with further tables for the d test of significance. *Ann. Eugen.* **11**, 141–172.

FISHER, R. A. and YATES, F. (1957) : *Statistical Tables*, Oliver and Boyd, Edinburgh, 5th ed., Table VI.

GOULDEN, C. H. (1952) : *Methods of Statistical Analysis*, John Wiley and Sons, New York, 2nd ed., 225.

Paper received : August, 1963.

A TOLERANCE REGION FOR MULTIVARIATE NORMAL DISTRIBUTIONS

By S. JOHN

Indian Statistical Institute

SUMMARY. In this paper is developed a method of determining from a random sample from a p-variate normal population a region regarding which it can be asserted that, with probability β, a proportion not less than α of the individuals in the population are contained in it. Solutions to some related problems are given in the final section.

1. Introduction

In most statistical populations, whether they be populations of income or of blood pressure or of tensile strength of metal castings, a preponderant majority of individuals are concentrated over a relatively narrow range. This enables us, in many situations, to act as if individuals falling outside such intervals did not exist. Theoretically, information concerning such regions is implicit in the probability distribution, though actual determination of them is often a matter of some mathematical difficulty, especially when the distribution is not unidimensional. The problem of tolerance regions is that of determining from only a random sample from the population, a region, regarding which we can assert that, with probability β, a proportion not less than α of the individuals in the population are contained in it.

The earliest formulation of the problem of tolerance regions is that of Wilks (1941). Wilks discovered a simple method of determining non-parametric tolerance regions for univariate populations. The corresponding multivariate problem was solved by Wald (1943). In Wald (1942) can be found an asymptotic solution of the problem of tolerance regions for parametric families of multivariate distributions. Tukey (1947, 1948), Tukey and Scheffé (1945), Fraser (1951, 1953), Fraser and Wormleighton (1951), Murphy (1948) and Kemperman (1956) report later work on the problem of non-parametric tolerance regions.

Though a general (asymptotic) solution of the problem of tolerance regions for a parametric family of distributions was given by Wald (1942), specialisation of his solution to particular families of distributions does not generally lead to the best or the simplest solution possible. On the other hand, the non-parametric solution can be inefficient, as demonstrated by Wilks (1941), when applied to such special families. For these reasons, Wald and Wolfowitz (1946) worked out a separate solution for univariate normal distributions. Our purpose in this paper is to work out such a solution for multivariate normal distributions.

2. NOTATION

We shall denote by $g_x(\mu, \Sigma)$ the density function of the multivariate normal distribution of dimension p, having μ for mean vector and Σ for dispersion matrix.

For any region R in the sample space, we shall set

$$\nu(R;\ \mu, \Sigma) = \int_R \cdots \int g_x(\mu, \Sigma)\, dx_1 dx_2 \ldots dx_p. \qquad \ldots \quad (2.1)$$

The function $f(z;\ \lambda, m)$ is defined as follows :

$$f(z;\lambda, m) = \frac{e^{-\lambda}}{(\sqrt{2})^m} \sum_0^\infty \frac{(\frac{1}{2}\lambda)^j}{j!\Gamma(\frac{1}{2}m+j)} z^{\frac{1}{2}m+j-1}\, e^{-\frac{1}{2}z}. \qquad \ldots \quad (2.2)$$

It is the density function of the non-central chi-square variable of noncentrality λ and degree of freedom m. Also, we shall set

$$F(z;\ \lambda,\ m) = \int_0^z f(u; \lambda,\ m)\, du. \qquad \ldots \quad (2.3)$$

The equation

$$F(P(\theta;\ \lambda,\ m); \lambda, m) = \theta \qquad \ldots \quad (2.4)$$

defines the function $P(\theta;\ \lambda, m)$.

3. PROCEDURE FOR DETERMINING THE TOLERANCE REGION

Let $\bar{x}$ be the arithmetic mean of N observations of a random vector $x = (x_1, \ldots, x_p)$ distributed according to the density function $g_x(\mu, \Sigma)$. Let V be a realisation of an independent Wishart variable of n degrees of freedom having $n\Sigma$ for its expectation.*

Denote by R_k the region of all x-vectors satisfying the inequality

$$(x-\bar{x})V^{-1}\ (x-\bar{x})' \leqslant k. \qquad \ldots \quad (3.1)$$

Let

$$v_1 = P(\alpha;\ \tfrac{1}{2}\, N^{-1}\, p, p), \qquad \ldots \quad (3.2)$$

and

$$v_2 = P(1-\beta;\ 0, n\, p). \qquad \ldots \quad (3.3)$$

Set

$$K = (v_1/v_2)p. \qquad \ldots \quad (3.4)$$

Regarding the region R_K, we can make the following assertion :

$$\text{prob}\ (\nu(R_K;\ \mu, \Sigma) \geqslant \alpha) \approx \beta. \qquad \ldots \quad (3.5)$$

The difference between the two members of (3.5) is small provided n and N are at least moderately large.

The constant v_2 can be determined from the table of the percentage points of the chi-square distribution given by Fisher and Yates (1953). If N is large $P(\alpha;\ N^{-1}p, p) \approx P(\alpha;\ 0,\ p)$. Hence when N is large, v_1 also can be determined from these same tables. If (p/N) is not small enough, we have to resort to methods developed by Patnaik (1949) and Abdel-Aty (1954). The short tables which they give would be of help in determining v_2.

---

* The matrix of corrected sum of products calculated from a random sample of size $n+1$ satisfies our requirements.

4. Proof of equation (3.5)

Let M be any non-singular matrix. Let $R_{k,M}$ be the region of all x-vectors satisfying the inequality

$$(x-\bar{x}M)(M'\,V\,M)^{-1}\,(x-\bar{x}M)' \leqslant k. \qquad \dots \quad (4.1)$$

Then,
$$\nu(R_k;\ \mu,\ \Sigma) = \nu(R_{k,M};\ \mu\,M,\ M'\Sigma\,M). \qquad \dots \quad (4.2)$$

We now choose A so that $A'\,\Sigma\,A = I$ and $A'\,V\,A$ is a diagonal matrix. Let the diagonal elements of $A'V\,A$ be $t_i (i = 1, 2, \dots, p)$. We can assume, without loss of generality, that $t_1 \leqslant t_2 \leqslant \dots \leqslant t_p$. The region $R_{k,A}$ is then the region of all x-vectors satisfying the inequality

$$\sum_1^p (x_i-a_i)^2/t_i \leqslant k, \qquad \dots \quad (4.3)$$

where a_i is the i-th component of the vector $\bar{x}\,A$.

Denote by R'_k the region of all x-vectors satisfying the inequality

$$\sum_1^p (x_i-a_i)^2 \leqslant k. \qquad \dots \quad (4.4)$$

Set
$$k' = (v_1/v_2)pt_1,\ k'' = (v_1/v_2)\sum_1^p t_i \qquad \dots \quad (4.5)$$

and
$$k''' = (v_1/v_2)pt_p. \qquad \dots \quad (4.6)$$

Equation (4.2), together with (4.3) and (3.4), leads to the following :

$$\text{prob}\,\{\nu(R'_{k'};\ \mu\,A, I) \geqslant \alpha\} \leqslant \text{prob}\,\{\nu(R_K;\ \mu, \Sigma) \geqslant \alpha\} \leqslant \text{prob}\,\{\nu(R'_{k'''};\ \mu A, I) \geqslant \alpha\}. \qquad \dots \quad (4.7)$$

Since $t_1 \leqslant (\Sigma\ t_i)/p \leqslant t_p$, we have also

$$\text{prob}\,\{\nu(R'_{k'};\ \mu A, I) \geqslant \alpha\} \leqslant \text{prob}\,\{\nu(R''_{k''};\ \mu A,\ I) \geqslant \alpha\} \leqslant \text{prob}\,\{\nu(R'_{k'''};\ \mu A,\ I) \geqslant \alpha\}. \qquad \dots \quad (4.8)$$

It is easy to demonstrate that

$$\text{prob}\,\{\nu(R'_{k'''}\,;\ \mu A,\ I) \geqslant \alpha\} - \text{prob}\,\{\nu(R'_{k'}\,; \mu A, I) \geqslant \alpha\}$$

tends to zero as $n \to \infty$. Therefore, *a fortiori*,

$$|\text{prob}\,\{\nu(R'_{k''};\ \mu A, I) \geqslant \alpha\} - \text{prob}\,\{\nu(R_K; \mu, \Sigma) \geqslant \alpha\}|$$

tends to zero as $n \to \infty$. Hence, equation (3.5) will be established if we show that

$$\text{prob}\,\{\nu(R'_{k''};\ \mu A, I) \geqslant \alpha\} \approx \beta. \qquad \dots \quad (4.9)$$

Now,
$$\nu(R'_k;\ \mu A,\ I) = F(k; w, p), \qquad \dots \quad (4.10)$$

where
$$w = \tfrac{1}{2}(\bar{x}-\mu)\Sigma^{-1}(\bar{x}-\mu)'. \qquad \dots \quad (4.11)$$

Therefore,

$$\begin{aligned}
&\text{prob}\,\{\nu\,(R'_{k''};\ \mu A,\ I) \geqslant \alpha\} \\
&\quad = \text{prob}\,\{F(k''; w, p) \geqslant \alpha\}, \\
&\quad = E_w\,\text{prob}\,\{v_1 v_2^{-1} v \geqslant P(\alpha; w, p)\,|\,w\},\ \text{where } v = \sum_1^p t_i, \\
&\quad = E_w\,\text{prob}\,\{v \geqslant v_2 v_1^{-1}\,P(\alpha; w, p)\,|\,w\}, \\
&\quad = 1 - E_w F(v_2 v_1^{-1}\,P(\alpha;\ w,\ p);\ 0,\ np), \\
&\quad = 1 - E_w F(v_2 v_1^{-1}\,P(\alpha; \tfrac{1}{2}N^{-1}p, p); 0, np) \\
&\qquad - E_w(w - \tfrac{1}{2}N^{-1}p)[(\partial/\partial\lambda)\,F(v_2 v_1^{-1} P(\alpha; \lambda, p); 0, np)]_{\lambda=\frac{1}{2}N^{-1}p} \\
&\qquad - \tfrac{1}{2}E_w(w - \tfrac{1}{2}N^{-1}p)^2[\partial^2/\partial\lambda^2)F(v_2 v_1^{-1}\,P(\alpha; \lambda, p); 0, np)]_{\lambda=\gamma(w)}, \quad \dots \quad (4.12)
\end{aligned}$$

by Taylor's theorem. Here $\gamma(w)$ is a function of w bounded by $\frac{1}{2}N^{-1}p$ and w.

Because of (3.2) and (3.3), the second term of the last member of (4.12) is $1-\beta$. The random variable $2Nw$ has the chi-square distribution with p degrees of freedom. From this it follows that the third term is zero. Finally, it is possible to prove that $(\partial^2/\partial\lambda^2)F(v_2 v_1^{-1}\,P(\alpha;\ \lambda,\ p);\ 0,\ np)$ is bounded. Further, $E(w-\frac{1}{2}N^{-1}p)^2 = \frac{1}{2}p/N^2$. Therefore, the absolute value of the fourth term is less than B/N^2, where B is some finite positive number. This proves that, if terms of order two in $(1/N)$ can be neglected, then

$$\text{prob}\,\{\nu(R'_{k''}; \mu A, I) \geqslant \alpha\} = \beta. \qquad \dots \quad (4.13)$$

5. Alternative procedures

Let $\xi(t_1, t_2, \dots, t_p)$ be any 'average' of $t_1, t_2, \dots, t_p$. Let v_3 be a number such that

$$\text{prob}\,\{\xi(t_1, t_2, \dots, t_p) < v_3\} = \beta. \qquad \dots \quad (5.1)$$

Set

$$k''' = v_1/v_3. \qquad \dots \quad (5.2)$$

We can then prove, by arguments exactly similar to those employed earlier, that

$$\text{prob}\,\{\nu(R_{k'''}\,;\ \mu,\ \Sigma) \geqslant \alpha\} \approx \beta. \qquad \dots \quad (5.3)$$

The procedure discussed in Section 3 corresponds to the choice of the arithmetic mean for ξ. This choice has the advantage that the exact value of v_3 can be determined quite easily using tables of the percentage points of the chi-square distribution. Some alternative choices for ξ are considered below.

(I) $\qquad \xi(t_1, t_2, \dots, t_p) = (t_1\, t_2, \dots, t_p)^{1/p}.$

Hoel (1937) shows that the density function of $\xi(t_1, t_2, \dots, t_p)$ is approximately

$$c^{\frac{1}{2}p(n-p+1)}\,\{\Gamma(\tfrac{1}{2}p[(n-p+1])\}^{-1}\;\xi^{\frac{1}{2}p(n-p+1)-1}e^{-c\xi}, \qquad \dots \quad (5.4)$$

where

$$c = \tfrac{1}{2}p[1-\tfrac{1}{2}(p-1)(p-2)/n]. \qquad \dots \quad (5.5)$$

If $p = 1$ or 2, (5.4) is the exact density function of $\xi(t_1, t_2, \dots, t_p)$. We can thus determine v_3 from tables of the chi-square distribution.

(II) $\qquad \xi(t_1, t_2, \dots, t_p) = p\Big/\Big(\sum_1^p t_i^{-1}\Big).$

In this case, $p\xi$ is distributed approximately as a chi-square with $\{np-p(p+1)+2\}$ degrees of freedom.

6. Bounds for prob $\{\nu(R_k; \mu, \Sigma) \geqslant \alpha\}$

If we neglect terms of order two in $(1/N)$,

$$\text{prob}\ (t_1 > v_2/p) \leqslant \text{prob}\ \{\nu(R_k;\ \mu, \Sigma) > \alpha\} \leqslant \text{prob}\ (t_p > v_2/p). \quad \dots \quad (6.1)$$

These inequalities are just another version of inequalities (4.7), obtained by the application of equation (4.10).

We give below simple expressions for the distribution functions of t_1 and t_r, in the case $p = 2$. Pillai (1954) gives recurrence relations connecting distribution functions of different orders.

If $p = 2$, starting from the joint distribution of t_1 and t_2, which is given, for instance, by Fisher (1939), we can show that

$$\text{prob}\ (t_1 > t) = [1-F(2t;\ 0, 2n)]-[\Gamma(\tfrac{1}{2})/\Gamma(\tfrac{1}{2}n)](t/2)^{\frac{1}{2}(n-1)}\, e^{-\frac{1}{2}t} \cdot [1-F(t;\, 0, n+1)] \quad \dots \quad (6.2)$$

and that,

$$\text{prob}\ (t_2 > t) = [1-F(2t;\ 0,\ 2n)]+[\Gamma(\tfrac{1}{2})/\Gamma(\tfrac{1}{2}n)](t/2)^{\frac{1}{2}(n-1)}\, e^{-\frac{1}{2}t}\, F(t,\ 0,\ n+1) \quad \dots \quad (6.3)$$

From (3.3), (6.2) and (6.3) we see that both extreme members of inequalities (6.1) are very nearly equal to β even for moderately large values of n. Therefore, the middle member is more so.

7. Related problems and concluding remarks

In some situations we face a slightly different problem. Here $\Sigma = \sigma^2\Lambda$ where Λ is a known positive definite matrix and σ^2 an unknown positive number. An unbiased estimate s^2 of σ^2, independent of the estimate $\bar{x}$ of μ, is available. The quantity ns^2 is a realisation of a chi-square variable with n degrees of freedom. A situation of this kind arises, for example, if we want to determine a tolerance region for the distribution of estimates of regression parameters in a linear model. The procedure of Section 3 applies to this case also if we set

$$ns^2\Lambda = V, \quad \dots \quad (7.1)$$

and

$$v_2 = P(\alpha;\ 0,\ n). \quad \dots \quad (7.2)$$

A problem closely related to that which we have been discussing in Sections 1 to 6 is that of determining from a random sample from the population a (random) region R regarding which we can make the following assertion :

$$E\nu(R;\, \mu,\, \Sigma) = \alpha. \quad \dots \quad (7.3)$$

The region R_k of Section 3 will satisfy this requirement if we choose k so that $kN(n-p+1)/[p(N+1)]$ is the upper $100(1-\alpha)$ percent point of the F-distribution with p and $n-p+1$ degrees of freedom. Fraser and Guttman (1956) prove that among regions satisfying condition (7.3), R_k is, in many respects, best.

In the practical application of the procedure of Section 3, it would be convenient to have at hand a table of values of K for various values of N, n and p. Such a table we hope to make available at a later date.*

* Tables required in the univariate case are given by Weissberg and Beatty (1960).

REFERENCES

ABDEL-ATY, S. H. (1954) : Approximate formula for the percentage points and the probability integral of the non-central χ^2 distribution. *Biometrika*, **41**, 538-540.

FISHER, R. A. (1939) : The sampling distribution of some statistics obtained from non-linear equations. *Ann. Eugen.*, **9**, 238-249.

FISHER, R. A. and YATES, F. (1953) : *Statistical Tables for Biological, Agricultural and Medical Research*, Fourth Edition, Oliver and Boyd, Edinburgh.

FRASER, D. A. S. (1951) : Sequentially determined statistically equivalent blocks. *Ann. Math. Stat.*, **22**, 294-298.

——— (1953) : Nonparametric tolerance regions. *Ann. Math. Stat.*, **25**, 44-55.

FRASER, D. A. S. and GUTTMAN, IRWIN (1956) : Tolerance regions. *Ann. Math. Stat.*, **27**, 162-179.

FRASER, D. A. S. and WORMLEIGHTON, R. (1951) : Nonparametric estimation, IV. *Ann. Math. Stat.*, **22**, 294-298.

HOEL, PAUL, G. (1937) : A significance test for component analysis. *Ann. Math. Stat.*, **8**, 149-158.

KEMPERMAN, J. H. B. (1956) : Generalized tolerance limits. *Ann. Math. Stat.*, **27**, 180-186.

MURPHY, R. B. (1948) : Nonparametric tolerance limits. *Ann. Math. Stat.*, **19**, 581-589.

PATNAIK, P. B. (1949) : The noncentral χ^2 and F distributions and their applications. *Biometrika*, **36**, 202-232.

PILLAI, K. C. S. (1954) : *On Some Distribution Problems in Multivariate Analysis*, Institute of Statistics Mimeograph Series No. 88, 34-39.

SCHEFFÉ, H. and TUKEY, J. W. (1945) : Nonparametric estimation, I. Validation of order statistics., *Ann. Math. Stat.*, **16**, 187-192.

TUKEY, J. W. (1947) : Nonparametric estimation, II. Statistically equivalent blocks and tolerance regions — the continuous case. *Ann. Math. Stat..*, **18**, 529-539.

——— (1948). Nonparametric estimation, III. Statistically equivalent blocks and multivariate tolerance regions — the discontinuous case. *Ann. Math. Stat.*, **19**, 30-39.

WALD, A. (1942) : Setting tolerance limits when the sample size is large. *Ann. Math. Stat.*, **13**, 389-399.

——— (1943) : An extension of Wilk's method of setting tolerance limits. *Ann. Math. Stat.*, **14**, 45-55.

WALD, A. and WOLFOWITZ, J. (1946) : Tolerance limits for a normal distribution. *Ann. Math. Stat.*, **17**, 208-215.

WEISSBERG, ALFRED and BEATTY, GLEN, H. (1960) : Tables of tolerance-limit factors for normal distributions. *Technometrics*, **3**, 483-501.

WILKS, S. S. (1941) : Determination of sample sizes for setting tolerance limits. *Ann. Math. Stat.*, **12**, 91-96.

——— (1942) : Statistical prediction with special reference to the problem of tolerance limits. *Ann. Math. Stat.*, **13**, 400-409.

Paper received : March, 1961.

ON TABLES OF RANDOM NUMBERS

By A. N. KOLMOGOROV
Academy of Sciences, USSR

1. Introduction

The set theoretic axioms of the calculus of probability, in formulating which I had the opportunity of playing some part (Kolmogorov, 1950), had solved the majority of formal difficulties in the construction of a mathematical apparatus which is useful for a very large number of applications of probabilistic methods, so successfully that the problem of finding the basis of real applications of the results of the mathematical theory of probability became rather secondary to many investigators.

I have already expressed the view [see Kolmogorov (1950), Chapter I] that the basis for the applicability of the results of the mathematical theory of probability to real 'random phenomena' must depend on some form of the *frequency concept of probability*, the unavoidable nature of which has been established by von Mises in a spirited manner. However, for a long time I had the following views.

(1) The frequency concept based on the notion of *limiting frequency* as the number of trials increases to infinity, does not contribute anything to substantiate the applicability of the results of probability theory to real practical problems where we have always to deal with a finite number of trials.

(2) The frequency concept applied to a large but finite number of trials does not admit a rigorous formal exposition within the framework of pure mathematics.

Accordingly I have sometimes put forward the frequency concept which involves the conscious use of certain not rigorously formal ideas about 'practical reliability', 'approximate stability of the frequency in a long series of trials', without the precise definition of the series which are 'sufficiently large' etc. [see *Foundations of the Theory of Probability*, Chapter I and for more details *Great Soviet Encyclopaedia* (section on Probability) and *Mathematika iou metod i Znachenye* (Chapter on Probability Theory)].

I still maintain the first of the two theses mentioned above. As regards the second, however, I have come to realise that the concept of random distribution of a property in a large finite population can have a strict formal mathematical exposition. In fact, we can show that in sufficiently large populations the distribution of the property may be such that the frequency of its occurrence will be almost the same for all sufficiently large sub-populations, when the *law of choosing these is sufficiently simple*. Such a conception in its full development requires the introduction of a measure of the complexity of the algorithm. I propose to discuss this question in another article. In the present article, however, I shall use the fact that there cannot be a *very large number of simple algorithms*.

For definiteness we shall consider the table

$$T = (t_1, t_2, \ldots, t_N)$$

of N zeros and ones : $\quad t_k = 0$ or 1.

Such a table will be called random, if, while choosing the subset A of sufficiently large size from $\overline{1, N}$ by different methods there is a stability in the frequency

$$\pi(A) = \frac{1}{n} \sum_{k \varepsilon A} t_k$$

of appearance of ones in A. One can, for example, choose A as

(a) the set of first n even integers $2, 4, 6, \ldots, 2n$

(b) the set of first n prime numbers $p_1, p_2, \ldots, p_n$ and so on.

The ordinary notion of 'randomness' of a table T does not consist merely of the stability of the frequencies while choosing A by methods entirely independent of the composition of the table T. One can for example, choose the set A as

(c) the set of first n values $k \geqslant 2$ for which $t_{k-1} = 0$,

(d) the set of first n values $k > s$ for which

$$t_{k-1} = a_1,\ t_{k-2} = a_2,\ \ldots,\ t_{k-s} = a_s,$$

(e) the set of the first n even numbers $k = 2i$ for which

$$t_i = 1,$$

(f) the set of numbers $k_1, k_2, \ldots, k_n \ldots$ chosen according to the law

$$k_1 = 1,$$

$$k_{i+1} = k_i + 1 + t_{k_i} p_i$$

and so on.

The precise formulation of the concept of 'admissible algorithm' of choosing the set A will be given in Section 2.

If while using a table of sufficiently large size N at least one single test of randomness of this type with sufficiently large size of the sample n leads to a 'significant' departure from the principle of frequency stability then we immediately reject the hypothesis of 'pure random' origin of the given table.

2. Admissible algorithms of selection and (n, ε)-random tables

An admissible algorithm of choosing the set

$$A = R(T) \subset \overline{1, N}$$

according to the table T of size N is defined by the functions*

$$F_0,\ G_0,\ H_0$$

$$F_1(\xi_1, \tau_1),\ G_1(\xi_1, \tau_1),\ H_1(\xi_1, \tau_1)$$

$$F_2(\xi_1, \tau_1;\ \xi_2, \tau_2),\ G_2(\xi_1, \tau_1;\ \xi_2, \tau_2),\ H_2(\xi_1, \tau_1;\ \xi_2, \tau_2)$$

$$\ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots$$

$$F_{N-1}(\xi_1, \tau_1;\ \xi_2, \tau_2;\ \ldots,\ \xi_{N-1}, \tau_{N-1}),\ G_{N-1}(\xi_1, \tau_1;\ \xi_2, \tau_2;\ \ldots;\ \xi_{N-1}, \tau_{N-1}),$$

$$H_{N-1}(\xi_1, \tau_1;\ \ldots;\ \xi_{N-1}, \tau_{N-1})$$

* The functions in the first line are constants (functions on the empty set of arguments).

where the arguments τ_k and the functions G_k and H_k take values 0 or 1 and the arguments ξ_k and functions F_k take values from $\overline{1, N}$. The functions F_k are subject to an additional condition

$$F_k(\xi_1, \tau_1; \ldots; \xi_k, \tau_k) \neq \xi_i. \qquad \ldots \quad (2.1)$$

Defining an algorithm is equivalent to forming the sequence

$$\begin{aligned} x_1 &= F_0 \\ x_2 &= F_1(x_1, t_{x_1};\ x_2\ t_{x_2}), \\ x_3 &= F_2(x_1, t_{x_1};\ x_2, t_{x_2}), \\ \ldots & \quad \ldots \quad \ldots \\ x_s &= F_{s-1}(x_1, t_{x_1}; \ldots; x_{s-1}, t_{x_{s-1}}) \end{aligned} \qquad \ldots \quad (2.2)$$

and determining those elements of the sequence which are found in A. The sequence terminates as soon as the value*

$$H_s(x_1, t_{x_1}; \ldots; x_s, t_{x_s}) = 1 \qquad \ldots \quad (2.3)$$

appears. In this case the sequence terminates with the element x_s. If when $k < N$ we have all the time

$$H_k(x_1, t_{x_1}; \ldots; x_k, t_{x_k}) = 0,$$

the sequence is terminated by the element x_s with $s = N$, i.e., by exhausting all the elements of the set $\overline{1, N}$: in view of the condition (2.1) all the elements of the sequence (2.2) are distinct.

The set A is formed from those x_k for which

$$G_{k-1}(x_1, t_{x_1}; \ldots; x_{k-1}, t_{x_{k-1}}) = 1 \qquad \ldots \quad (2.4)$$

It seems to me that the given construction correctly reflects the basic concept of von Mises in its complete generality, preserving, however, the basic limitation that for determining whether $x \epsilon \overline{1, N}$ falls in the set A the value of t_x is not used.

Now let the system

$$\mathcal{R}_N = \{R\}$$

of admissible algorithms of selection (the size N of the table being fixed) be given.

Definition : The table T of size N is called (n, ε)-random with respect to the system $\mathcal{R}_N$, if there exists a constant p, $0 \leqslant p \leqslant 1$, such that for any

$$A = R(T),\ R \epsilon \mathcal{R}_N$$

with the number of elements $\quad V \geqslant n,$

the frequency $\quad \pi(A) = \frac{1}{v} \sum_{k \varepsilon A} t_k$

satisfies the inequality $\quad |\pi(A) - p| \leqslant \varepsilon.$

---

* In particular, if $H_0 = 1$ then the selection cannot begin and the set A is found to be empty.

Sometimes, it is convenient to say (n, ε, p)-randomness, assuming that the constant p is fixed. Then the following theorem holds.

Theorem 1 : *If the number of elements of the system $\mathcal{R}_N$ does not exceed*

$$\tau(n, \varepsilon) = \frac{1}{2} e^{2n\varepsilon^2} \qquad \dots \quad (2.5)$$

then for any p, $0 \leqslant p \leqslant 1$, there exists a table T of size N that is (n, ε, p)-random with respect to $\mathcal{R}_N$.

The interpretation of the estimate, contained in the theorem, is made more transparent, if we introduce the binary logarithm

$$\lambda(\mathcal{R}_N) = \log_2 \rho(\mathcal{R}_N)$$

of the number of elements ρ of the system $\mathcal{R}_N$. $\lambda(\mathcal{R}_N)$ is equal to the quantity of information, which is necessary for choosing an individual element R from $\mathcal{R}_N$. It is clear that in the case of large $\lambda(\mathcal{R}_N)$ the system $\mathcal{R}_N$ must contain algorithms, the very determination (and not merely the actual realisation) of which is complicated (requires for its formulation not less than $\lambda(\mathcal{R}_N)$ binary symbols).

In our theorem the condition of existence of tables which are (n, ε)-random with respect to $\mathcal{R}_N$ with arbitrary p is written in the form of the inequality

$$\lambda(\mathcal{R}_N) \leqslant 2 \log_2 en\varepsilon^2(1-\varepsilon)-1. \qquad \dots \quad (2.6)$$

Such a qualitative formulation of the result contained in the theorem is instructive by itself. If the ratio λ/n is sufficiently small then for any previously given ε and any N and p there exist tables which are (n, ε)-random with respect to any system of admissible algorithms with

$$\lambda(\mathcal{R}_N) \leqslant \lambda.$$

The proof of this theorem will be given in Section 3. In Section 4, we shall examine the possibility of improving the estimates contained in the theorem. Now we make two supplementary remarks.

Remark 1 : Since the algorithm of choosing the set $A = R(T)$ is determined by the functions F_k, H_k, G_k it is natural to consider two algorithms to be same when and only when their corresponding functions F_k, H_k, G_k coincide. Already from this point of view the number of distinct possible algorithms of selection for a given N is finite.

It is possible to hold on to a different point of view and consider two algorithms of selection to be different only in the case when they give different sets $A = R(T)$ at least for one table T. From such a point of view the number of distinct algorithms is further reduced. But in any case it is not greater than

$$(2^N)^{2^N} = 2^{N.2^N}.$$

The question of precise estimation of the number of admissible algorithms under the second approach is not so simple. The problem is very simple only for algorithms, by which the set A is formed independently of the properties of the table T. Distinct number of such algorithms is equal to 2^N according to the number of different sets $A \epsilon \overline{1, N}$.

Remark 2 : The admissible algorithms of selection from the set of all possible natural numbers was considered by Church (1940). Now, in our definition, instead of the finite table T we consider an infinite sequence of zeroes and ones,

$$t_1, t_2, \ldots \quad .$$

We assume that the values of the arguments ξ_k and the function F_k are arbitrary natural numbers. But we reject the requirement that the selection must stop at $s = N$ and instead assume that any (now infinite) table of functions F_k, G_k, H_k is 'computable' in the sense sufficiently well-known in all the numerous propositions for such formal definitions. Under these considerations we obtain the inessential generalisation of Church's concept. The basis of Church's result is the existence, for any p, of sequences $t_1, t_2, \ldots, t_N \ldots$the density of which is equal to p in any infinite[1] set A obtained by an admissible algorithm.

3. Proof of theorem 2

This result belongs to the Theory of Finite Algorithms and its formulation does not contain any concept borrowed from Probability Theory. If, in proving this, we make use of certain results of Probability Theory then this proof will have a formal character as it would only include a certain distribution of 'weights' in the set of tables T of size N, the weight

$$P(T) = p^M(1-p)^{N-M}$$

being assigned to the table containing M ones. This method of proof does not affect the logical nature of the theorem itself, and does not hinder its use in the discussions needed for defining the domain of applicability of Probability Theory.

In another paper we shall prove the following inequality relating to the 'Bernoulli Scheme' :

$$P\left(\sup_{k \geqslant n} \left| \frac{\mu_k}{k} - p \right| \leqslant \varepsilon \right) \leqslant 2\ e^{2n\varepsilon^2(1-\varepsilon)} \qquad \ldots \quad (3.1)$$

Here p is the probability of success in each of a sequence of independent trials; μ_k is the number of successes in the first k trials. We can easily derive the following corollary from (3.1).

Corollary : *Let*[2]

$$P(\xi_k = 1 |\ k \leqslant \nu,\ \xi_1, \ldots, \xi_{k-1}) = p$$

where $\xi_1, \xi_2, \ldots, \xi_\nu$ *is a sequence of a random number of random quantities and* p *is a constant. Then*

$$P\left(\nu \geqslant n, \left| \frac{\mu_\nu}{\nu} - p \right| \geqslant \varepsilon \right) \leqslant 2\ e^{-2n\varepsilon^2(1-\varepsilon)}. \qquad \ldots \quad (3.2)$$

---

[1] In this concept of Church substantial interest lies only in algorithms which extend infinitely. That is why, in this case, the functions G_k and all that is connected with these functions must be omitted.

[2] We are concerned here with the conditional probability that $\xi_k = 1$ when $k = \nu$ and ξ_1, ξ_2, ξ_{k-1} are given.

We shall now examine the system $\mathcal{R}_N$ of admissible algorithms, ρ in number.

We consider a table formed randomly with probability p for $t_x = 1$ independently of the values taken by the other $t_{m'}$. If we fix $R \epsilon \mathcal{R}_N$ and denote by

$$\xi_1, \xi_2, \ldots, \xi_\nu$$

those elements of the sequence

$$x_1, x_2, \ldots, x_s$$

which fall in $A = R(T)$ (numbering them as they appear in the course of the algorithm) it can easily be seen that the conditions under which (3.2) is valid are fulfilled. Hence the probability that, for any given $R \epsilon \mathcal{R}_N$, the number of elements ν of the set A will not be less than n and the inequality $|\pi(A)-p| \geqslant \varepsilon$ will also be satisfied, will be less than $2\, e^{-2n\varepsilon^2(1-\varepsilon)}$.

If

$$\rho \leqslant \frac{1}{2}\, e^{2n\varepsilon^2(1-\varepsilon)}$$

then the sum of the probabilities of failure of the inequality

$$|\pi(A)-p| \leqslant \varepsilon$$

for those algorithms which lead to the sets with not less than n elements will be less than unity. Hence with positive probability the table T will be found to be (n, ε, p)—random in the sense of the definition of Section 2. Hence follows the existence of tables which are (n, ε, p)-random with respect to $\mathcal{R}_N$ (indeed independently of the probabilistic assumptions on the distribution of $P(T)$ in the space of tables).

4. On the possibilities of improving the estimate by the theorem of Section 2

If we fix n, ε, N, p, then, for an integral non-negative ρ one of the two situations is possible :

(a) whatever be the system $\mathcal{R}_N$ of ρ admissible algorithms of selection, there exists a table T of size N which is (n, ε, p)-random with respect to $\mathcal{R}_N$;

(b) there exists a system $\mathcal{R}_N$ of ρ admissible algorithms of selection relative to which there are no (n, ε, p)-random tables T of size N.

We can easily find that the existence of the situation (a) for some ρ follows from the existence of the same situation for $\rho' < \rho$. It is clear that for $\rho = 0$ the situation (a) will always be true. Hence, there exists an upper bound

$$\tau(n, \varepsilon, N, p) = \sup_{\rho\, \varepsilon\, a} \rho$$

of those ρ for which the case (a) holds. For all ρ greater than $\tau\,(n, \varepsilon, N, p)$ the case (b) holds.

If we put

$$\tau\,(n, \varepsilon) = \inf_{p,\, N} \tau(n, \varepsilon, N, p)$$

then the substance of the theorem of Section 2 can be expressed in the form of the inequality:

$$\tau(n, \varepsilon) \geqslant \frac{1}{2} e^{2n\varepsilon^2(1-\varepsilon)}. \qquad \dots (4.1)$$

Now taking logarithms

$$l(n, \varepsilon, N, p) = \log_2 \tau(n, \varepsilon, N, p), \quad l(n, \varepsilon) = \log_2 \tau(n, \varepsilon),$$

we can write (2.6) in the form

$$l(n, \varepsilon) \geqslant 2n\varepsilon^2(1-\varepsilon)-1. \qquad \dots (4.2)$$

In fact, the main interest lies in the asymptotically precise estimation of $l(n, \varepsilon)$ when ε is small and n and $l(n, \varepsilon)$ are large. When

$$\varepsilon \to 0, \; n\, \varepsilon^2 \to \infty$$

we get from (4.2)

$$l(n, \varepsilon) \geqslant 2n\varepsilon^2+o(n\varepsilon^2). \qquad \dots (4.3)$$

We shall find later, on the other hand, that when

$$\varepsilon \to 0, \; n\varepsilon \to \infty$$

the relation

$$l(n, \varepsilon) \leqslant 4n\varepsilon+o(n\varepsilon) \qquad \dots (4.4)$$

will hold. Unfortunately, I cannot remove the discrepancy between the power of ε in (4.3) and (4.4).

The estimate (4.4) is a simple consequence of the following theorem the formulation of which is unfortunately somewhat complex and will become clear through the method of proof chosen by us.

Theorem 2 : *If* $k \leqslant \dfrac{1-2\varepsilon}{4\varepsilon}$, $n \leqslant (k-1)m$, $N \geqslant km$ *then*

$$\tau(n, \varepsilon, N, \tfrac{1}{2}) \leqslant k.2^m. \qquad \dots (4.5)$$

For proving the theorem it is enough to construct, under the condition

$$k \leqslant \frac{1-2\varepsilon}{4\varepsilon}, \; n = (k-1)m, \quad N = km,$$

a system $\mathcal{R}_N$ out of

$$\rho = k.2^m+1$$

admissible algorithms, for which there does not exist an $(n, \varepsilon, \frac{1}{2})$-random table T.

We partition $\overline{1, N}$ into k sets Δ_i, $i = 1, \dots k$, with m elements in each. Every Δ_i contains 2^m subsets. We form the set

$$A_{is}, \quad i = 1, 2, \dots, k; \quad s = 1, 2, \dots, 2^m$$

by taking the union of all Δ_j, $j \neq i$ and the ρ-th subset of Δ_i. We form the system $\mathcal{R}_N$ from

(a) $k.2^m$ algorithms R_{i_s} for selecting the sets A_{is};

(b) one algorithm R for selecting $A = \overline{1,N}$.

We prove that there does not exist a table T which is $(n, \varepsilon, \frac{1}{2})$-random with respect to R_N.

Let us take an arbitrary table T and assume that it is $(n, \varepsilon, \frac{1}{2})$-random with respect to R_N. Then it must contain at least $(\frac{1}{2}-\varepsilon)N$ zeroes and $(\frac{1}{2}-\varepsilon)N$ ones. Hence we can find i and j such that Δ_i contains $\alpha \geqslant (\frac{1}{2}-\varepsilon)m$ zeroes and Δ_j contains $\beta \geqslant (\frac{1}{2}-\varepsilon)m$ ones.

Let
$$\gamma = \min(\alpha, \beta) \geqslant (\tfrac{1}{2}-\varepsilon)m.$$

There exists an algorithm $R' \epsilon R_N$ $(R'' \epsilon R_N)$ for selecting the set A' (A'') which consists of the entire $\overline{1, N}$ except γ elements in $\Delta_i(\Delta_j)$ which correspond to zero (one) in the table T. It is easy to see that the corresponding frequencies are equal to

$$\pi(A') = \frac{M}{N-\gamma}, \; \pi(A'') = \frac{M-\gamma}{N-\gamma}$$

where M is the total number of ones in the table T. Let us estimate the difference between these frequencies :

$$\lambda(A')-\lambda(A'') = \frac{\gamma}{N-\gamma} \geqslant \frac{(\frac{1}{2}-\varepsilon)m}{km} \geqslant 2\varepsilon.$$

This estimate contradicts the set of inequalities

$$|\pi(A')-\tfrac{1}{2}| \leqslant \varepsilon, |\pi(A'')-\tfrac{1}{2}| \leqslant \varepsilon,$$

which follow from the hypothesis of $(n, \varepsilon, \frac{1}{2})$-randomness of the table T. This contradiction proves the theorem.

REFERENCES

KOLMOGOROV, A. N. (1950) : *Foundations of the Theory of Probability*, Chelsea.

——— Section on 'Probability', *Great Soviet Encyclopaedia*, 2nd edition.

——— Chapter on Probability Theory in the book *Matematika iou metod i Znachenye*, Academy of Sciences, USSR.

CHURCH, (1940) : On the concept of a random sequence. *Bull. Amer. Math. Soc.*, **46**, 130–135.

Paper received : April, 1963.

REMARKS ON THE BEHRENS-FISHER PROBLEM

By YU. V. LINNIK

Mathematical Institute, Leningrad

1

The Behrens-Fisher problem is well known in the statistical literature (see for references J. Neyman's report at the Amsterdam Mathematical Congress, 1954). We are given two independent samples

$$x_1, \ldots, x_n \in N(a_1, \sigma_1);\ y_1, \ldots, y_n \in N(a_2, \sigma_2)$$

consisting of independent identically distributed observations belonging to two normal populations with the parameters a_1, σ_1; a_2, σ_2. The general Behrens-Fisher problem consists in constructing the tests for the hypothesis $H_0 : a_1 = a_2$ when nothing is known about the ratio.

We consider the tests based upon the similar zones excluding the nuisance parameters σ_1, and σ_2. They will be determined by the statistics t which we shall call similar statistics. These are the statistics t measurable with respect to all the probability measures induced by $N(a_1, \sigma_1)$ and $N(a_2, \sigma_2)$; for a similar statistic we have by definition : $p(t < \xi)$ for any ξ does not depend upon σ_1 and σ_2 on the hypothesis $H_0 : a_1 = a_2$.

It is well known that there are non-trivial similar statistics (Romanovsky-Bartlett-Scheffé tests) (see Barankin, 1950). The system of sufficient statistics is formed by $\bar{x}, \bar{y}, s_1^2, s_2^2$. It is well known that the Neyman structures (see Lehmann, 1959) in the case of this particular problem produce similar statistics t indeed, but these are useless because $p(t < \xi)$ does not depend upon all the parameters of the problem and so cannot discern H_0 from the alternatives.

If $\sigma_1 = \sigma_2$, we obtain the classical problem of Student. Student's similar statistic excluding $\sigma_1 = \sigma_2$, is well known as :

$$t_1 = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{m}{n}\, s_1^2 + s_2^2}}. \qquad \ldots \quad (1.1)$$

We remark that this statistic depends explicitly only upon $\bar{x}-\bar{y}$, s_1^2, s_2^2 and the ratio of the sample sizes $\frac{m}{n}$. For Student's problem the system of sufficient and necessary statistics is $\left\{ \sum_{i=1}^{m} x_i + \sum_{j=1}^{n} y_j ; \sum_{i=1}^{m} x_i^2 + \sum_{j=1}^{n} y_j^2 \right\}$

and t_1 is independent of them thus forming a Neyman structure. One of the aspects of the Behrens-Fisher problem consists in the investigation of the similar statistics which are dependent only upon the sufficient statistics $\bar{x}, \bar{y}, s_1, s_2$ (and so cannot be nontrivial Neyman structures). They are usually considered to be more all less formally similar to the Student statistic t_1. (see Wald, 1955).

We shall consider here the similar statistics for the Behrens-Fisher problem which are in a certain respect similar to the Student statistic t_1. Namely, we shall consider the statistic t with the following properties.

(I) $$t = t\left(\bar{x}-\bar{y},\ s_1^2,\ s_2^2,\ \frac{m}{n}\right)$$

which, depends only upon the arguments $\bar{x}-\bar{y},\ s_1^2, s_2^2, \frac{m}{n}$. The sample sizes enter into t only implicitly and in the form of the ratio $\frac{m}{n}$.

Denote $\bar{x}-\bar{y} = X$; $s_1^2 = u$; $s_2^2 = v$. Let $\xi > 0$ be any positive number. We require that

(II) $$t\left(\xi X,\ \xi^2 u,\ \xi^2 v, \frac{m}{n}\right) = t\left(X, u, v, \frac{m}{n}\right)$$

(homogeneity property).

In view of the property (II), we have for any $v > 0$:

$$t\left(X, u, v, \frac{m}{n}\right) = t\left(\frac{X}{\sqrt{v}}, \frac{u}{v}, 1, \frac{m}{n}\right) = g\left(\frac{X}{\sqrt{v}}, \frac{u}{v}, \frac{m}{n}\right) \quad \text{say.}$$

We introduce the property

(III) $g\left(\frac{X}{\sqrt{v}}, \frac{u}{v}, \frac{m}{n}\right)$ for the fixed $\frac{m}{n}$ which is a continuous function in both variables

$$\xi = \frac{X}{\sqrt{v}},\ \xi = \frac{u}{v} \quad \text{for } -\infty < \xi < \infty;\ 0 \leqslant \xi < \infty.$$

The Student's statistic t_1 (c.f. (1. 1)) will obviously satisfy the requirements (I), (II), and (III).

2

We formulate now the two theorems which are to be proved in this paper.

Theorem 1 : *A non-trivial similar statistic z for the Behrens-Fisher problem cannot be continuous. It must have the discontinuity points at least at each point of the set*

$$x_1 = x_2 = \ldots = x_m = y_1 = \ldots y_n.$$

Theorem 2 : *Any similar statistic* $t = t\left(\frac{X}{\sqrt{v}}, \frac{u}{v}, \frac{m}{n}\right)$ *for the Behrens-Fisher problem, satisfying the requirements* (I), (II) *and* (III) *must be constant for* $X = 0$: *and* $t\left(0, \frac{u}{v}, \frac{m}{n}\right)$ *does not depend upon* $\frac{u}{v}$.

Though these theorems are of a special type, the methods of proof might be of some more general interest.

3

We now pass to the proof of Theorem 1. Let z be a non-trivial similar statistic for the Behrens-Fisher problem. Put $a_1 = a_2 = a$. Consider a as a given and fixed number, not a parameter. The likelihood function will be obviously:

$$L(x_1, \ldots, x_m, y_1 \ldots, y_n) = (2\pi)^{-(m+n)/2} \sigma_1^{-m} \ \sigma_2^{-n}.$$

$$\exp\left(-\frac{1}{2\sigma_1^2} \sum_{i=1}^{m} (x_i-a)^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^{n} (y_i-a)^2\right).$$

Now a being a fixed number we see that $\sum_{i=1}^{m} (x_i-a)^2$ and $\sum_{j=1}^{n} (y_i-a)^2$ are sufficient statistics for the parameters $\frac{1}{\sigma_1^2}$ and $\frac{1}{\sigma_2^2}$. It is obvious that the conditions of the Lehmann-Scheffé theorem on complete systems of sufficient statistics are satisfied (cf. Lehmann (1954), p. 132, Theorem 1). Hence z must be a Neyman structure and thus independent of $\left(\sum_{i=1}^{m} (x_i-a)^2, \sum_{j=1}^{n} (y_i-a)^2\right)$. Suppose, $z(x_1, \ldots, x_m, y_1, \ldots, y_n)$ to be continuous at the point $(a, a, \ldots, a)$, but non-trivial. Hence, there exist $\delta > 0$ and $\varepsilon > 0$ such that

$$p\{|z(x_1, \ldots, x_m; y_1, \ldots, y_n)-z(a, a, \ldots, a)| > \delta\} > \varepsilon. \qquad \ldots \quad (3.1)$$

But z must be independent of the pair

$$\left(\sum_{i=1}^{m} (x_i-a)^2, \sum_{j=1}^{n} (y_j-a)^2\right).$$

If we put $$\sum_{i=1}^{m} (x_i-a)^2 = \varepsilon_1; \quad \sum_{j=1}^{n} (y_j-a)^2 = \varepsilon_1,$$

then for a sufficiently small ε_1, $|x_i-a|$ and $|y_j-a|$ will be sufficiently small ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$). But then we shall have :

$$p\{|z(x_1, \ldots, x_m, y_1, \ldots, y_n)-z(a_1, \ldots, a)| < \delta\} = 1 \qquad \ldots \quad (3.2)$$

which contradicts (3.1).

As this is true for any point $x_1 = \ldots = x_m = y_1 = \ldots = y_n = a$ our assertion is proved.

4

Let now $$t = t\left(X, u, v, \frac{m}{n}\right) = t_0\left(\frac{X}{\sqrt{v}}, \frac{u}{v}, \frac{m}{n}\right)$$

be a similar statistic for the Behrens-Fisher problem satisfying the requirements (I), (II) and (III); we have :

$$p(t_0 < \xi) \qquad \ldots \quad (4.1)$$

which does not depend upon σ_1 and σ_2 if $a_1 = a_2$.

Denote $$\sigma_1^2 = \theta_1 \, ; \ \sigma_2^2 = \theta_2 \, ; \ \frac{m}{n} = \frac{1}{v},$$

and put $$\frac{X}{(v\theta_1+\theta_2)^{\frac{1}{2}}} = X_1 \, ; \quad \frac{u}{\theta_1} = u_1 \, ; \quad \frac{v}{\theta_2} = v_1.$$

Then $X_1 \epsilon N\left(0, \frac{1}{\sqrt{n}}\right)$ is a normal variable, while $u_1 = \frac{1}{n_1} \chi^2_{n-1}$; $v_1 = \frac{1}{n} \chi'^2_{n-1}$ are χ^2-type variables; X_1, u_1, v_1, are stochastically independent. Hence, we have :

$$t\left(X, u, v, \frac{m}{n}\right) = t\left(\frac{X_1}{\sqrt{v_1}}(1+v\theta)^{\frac{1}{2}}, \frac{u_1}{v_1}\theta, 1, \frac{m}{n}\right) \quad \text{where } \theta = \frac{\theta_1}{\theta_2},$$

which must have the distribution independent of θ, for any $\theta > 0$. The same property must hold for any continuous function $\psi(t)$. Hence it follows easily that the Lebesgue sets : $\underset{x_1, u_1, v_1}{\varepsilon}(t < \lambda)$ for any λ must have the measure induced by X_1, u_1, v_1 independent upon θ (this is also sufficient for t to be a similar statistic). Let now $n \to \infty$, the ratio $\frac{m}{n}$, remaining constant. Then the probability measure induced by $\frac{X}{\sqrt{v_1}}, \frac{u_1}{v_1}$ obviously becomes concentrated around the point :

$$\xi = \frac{X}{\sqrt{v_1}} = 0; \quad \eta = \frac{u_1}{v_1} = 1\,.$$

Now let the value of t in the point $\xi = 0$, $\eta = 1$ for $\theta = 1$ be : $t\left(0, 1, 1, \frac{m}{n}\right) = t_1$, say. For a fixed arbitrarily small ε, consider the Lebesgue set:

$$\varepsilon(t_1 - \varepsilon \leqslant t < t_1 + \varepsilon).$$

For $n \to \infty$ the measure of this set, induced by X_1, u_1, v_1 will converge to 1. Now, for a given value of $\theta > 0$, we have :

$$t\left(0, \theta, 1, \frac{m}{n}\right) = t_\theta \neq t_1.$$

As t is continuous, for sufficiently small ε, we shall have :

$$p\left(t_1 - \varepsilon < t\left(0, \theta, 1, \frac{m}{n}\right) < t_1 + \varepsilon\right) \to 0 \qquad \text{for } n \to \infty.$$

But this probability must not depend upon the value of θ. Hence $t_\theta = t_1$, and $t\left(0, \theta, 1, \frac{m}{n}\right)$ must be a constant.

REFERENCES

NEYMAN, J. (1954) : Current problems of mathematical statistics. Report at the International Congress of Mathematicians, Amsterdam, 1954.

BARANKIN, E. W. (1950) : Extension of the Romanovsky-Bartlett-Scheffé Test. *Proceedings of the First Berkeley Symposium Mathematical Statistics and Probability*, 433-450.

LEHMANN, E. L. (1959) : *Testing Statistical Hypothesis*, John Wiley & Sons, New York.

WALD, A. (1955) : *Selected Papers in Statistics and Probability*, Stanford University Press, California.

Paper received : April, 1963.

A NOTE ON DETERMINATION OF SAMPLE SIZE

By M. N. MURTHY

Indian Statistical Institute

SUMMARY. In this note a procedure of determining the sample size is proposed, where the idea is to fix the sample size in such a way that the probability (P) of the length (L) of the confidence interval (associated with a specified confidence coefficient) for the parameter (μ) being less than a given value ($k\mu$) is a pre-specified quantity.

Let y be normally distributed with mean μ and standard deviation σ. Suppose a sample of N units is drawn with equal probability. Then the mean $\bar{y}$ based on the N observations is normally distributed with mean μ and standard error $\sigma/\sqrt{N}$. Let s^2 be an unbiased estimator of σ^2 based on a sub-sample of n units (or on n random sub-groups of $\frac{N}{n}$ units),

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

It is well known that the statistic

$$t = \frac{\bar{y}-\mu}{s/\sqrt{N}}$$

is distributed as Student's t with $(n-1)$ degrees of freedom.

Using the tabulated values of the t-distribution, we can set up confidence interval for μ at any specified level of confidence $(1-\alpha)$. That is, if t_α is the $\alpha\%$ point of t, then

$$P\left[\frac{|\bar{y}-\mu|}{s/\sqrt{N}} < t_\alpha\right] = 1-\alpha. \qquad \ldots \quad (1)$$

The length L of the confidence interval is given by

$$L = 2t_\alpha s/\sqrt{N}. \qquad \ldots \quad (2)$$

Suppose the sample size is to be so fixed that

$$P(L < k\mu) = 1-\beta, \qquad \ldots \quad (3)$$

where k is a pre-specified quantity and $(1-\beta)$ may be taken as the *second level of confidence*, the first level of confidence being $(1-\alpha)$ in (1). It may be noted that $P(L < k\mu)$ is a function of the sample size and increases with increase in sample size.

For finding the sample size which could satisfy both the levels of confidence given in (1) and (3), we may proceed as follows.

$$P(L < k\mu) = 1-\beta,$$

that is, $$P(2t_\alpha s/\sqrt{N} < k\mu) = 1-\beta,$$

that is, $$P\left[\frac{(n-1)s^2}{\sigma^2} < \frac{k^2}{c^2}\frac{N(n-1)}{4t_\alpha^2}\right] = 1-\beta, \quad \ldots \quad (4)$$

where c is the population coefficient of variation (σ/μ) and $(n-1)s^2/\sigma^2$ is a χ^2 with $(n-1)$ degrees of freedom. Reducing (4) to an incomplete Γ-function which is already tabulated, we get

$$P(L < k\mu) = I(u, p) \quad \ldots \quad (5)$$

where $$I(u, p) = \frac{1}{\Gamma(p+1)} \int_0^{u\sqrt{(p+1)}} e^{-x}\, x^p dx,$$

$$p = (n-1)/2, \text{ and } u = \frac{k^2}{c^2}\frac{N\sqrt{(n-1)}}{4\sqrt{2}t_\alpha^2}.$$

For given values of $(1-\alpha)$, $(1-\beta)$, c, n, and k we can first get the value of u such that

$$I(u, p) = 1-\beta$$

and then get the required sample size

$$N = u\frac{c^2}{k^2}\frac{4\sqrt{2}\,t_\alpha^2}{\sqrt{(n-1)}}. \quad \ldots \quad (6)$$

The usual procedure of determining the sample size consisted in finding the value of N such that $E(L)$ is equal to a specified value $k\mu$. The proposed procedure given in this note is a generalization of the usual procedure in the sense that it ensures a pre-specified value for the probability that L is less than $k\mu$.

Paper received : February, 1962.

ON SPECTRAL ANALYSIS WITH MISSING OBSERVATIONS AND AMPLITUDE MODULATION*

By EMANUEL PARZEN

Stanford University

SUMMARY. The notion of an asymptotically stationary time series and its spectral analysis was considered by the author (Parzen, 1961b). An important example of an asymptotically stationary time series is an amplitude modulated stationary time series. In this note, the problem of spectral analysis of stationary normal time series with missing observations, recently treated by Jones (1962), is treated as a special case of the problem of spectral analysis of an amplitude modulated stationary normal time series.

1. Introduction

Let $\{X(t),\ t = 1, 2, \ldots\}$ be a discrete parameter time series with zero means and finite second moments. It is said to be *weakly* (see Doob, 1953) *or covariance* (see Parzen, 1962) *stationary* if there exists a function, denoted $R(v)$ and called the covariance function of the time series, such that for $v = 0, 1, 2, \ldots,$

$$R(v) = E[X(t)\, X(t+v)] \qquad \ldots \quad (1.1)$$

independently of $t = 1, 2, \ldots$. It is said to be asymptotically (weakly) stationary if instead of (1.1) it holds that

$$R(v) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-v} E[X(t)\, X(t+v)]. \qquad \ldots \quad (1.2)$$

If either (1.1) or (1.2) hold, the time series is said to be *ergodic* if the sample covariance function

$$R_T(v) = \frac{1}{T} \sum_{t=1}^{T-v} X(t)\, X(t+v) \qquad \ldots \quad (1.3)$$

is, for $v = 0, 1, \ldots,$ a consistent in quadratic mean estimate of $R(v)$. In order for this to be the case it is necessary and sufficient that for each v

$$\lim_{T\to\infty} \operatorname{var}[R_T(v)] = 0. \qquad \ldots \quad (1.4)$$

One important way in which asymptotically stationary time series arise is by *amplitude modulating a (covariance) stationary process.*

Let $\{Y(t),\ t = 1, 2, \ldots\}$ be a stationary time series with zero means and covariance function

$$R_Y(v) = E[Y(t)\, Y(t+v)]. \qquad \ldots \quad (1.5)$$

*Prepared under Contract Nonr-225 (21), (NR-042-993), for Office of Naval Research; reproduction in whole or in part is permitted for any purpose of the United States Government.

Let $\{g(t), t = 1, 2, \ldots\}$ be a non-random bounded function possessing a generalized harmonic analysis in the sense that for $v = 0, 1, \ldots$

$$R_g(v) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T-v} g(t)\, g(t+v) \qquad \ldots \quad (1.6)$$

exists. The time series

$$X(t) = g(t)\, Y(t) \qquad \ldots \quad (1.7)$$

may be called the original time series $Y(\bullet)$ amplitude modulated by the function $g(\bullet)$. Since

$$E[X(t)\, X(t+v)] = g(t)\, g(t+v)\, R_Y(v) \qquad \ldots \quad (1.8)$$

it is clear that while $X(\bullet)$ is not covariance stationary, it is asymptotically stationary with covariance function $R_X(v)$ given by

$$R_X(v) = R_g(v)\, R_Y(v). \qquad \ldots \quad (1.9)$$

It is shown by the author (Parzen, 1961b) that if $Y(\bullet)$ is an ergodic normal process, then $X(\bullet)$ is ergodic. Consequently, given observations $\{X(t), t=1, 2, \ldots, T\}$ a consistent (in quadratic mean) estimate of $R_X(v)$ is given by the sample covariance function

$$R_T(v) = \frac{1}{T} \sum_{t=1}^{T} X(t)\, X(t+v). \qquad \ldots \quad (1.10)$$

A consistent estimate of $R_Y(v)$ is then available at all lags v for which $R_g(v) \neq 0$, namely

$$\hat{R}_Y(v) = R_T(v)/R_g(v). \qquad \ldots \quad (1.11)$$

From these facts we obtain immediately the following theorem.

Theorem 1A : *Let $\{Y(t),\ t = 1, 2, \ldots\}$ be stationary and normal with zero means and covariance function $R_Y(v)$ satisfying*

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} R^2(v) = 0, \qquad \ldots \quad (1.12)$$

so that $Y(\bullet)$ is ergodic.

Suppose that the time series $Y(\bullet)$ is not directly observed. Rather one observes a time series $X(\bullet)$ which is an amplitude modulated version of $Y(\bullet)$:

$$X(t) = g(t)\, Y(t), \qquad t = 1, 2, \ldots, \qquad \ldots \quad (1.13)$$

where $g(\bullet)$ is a non-random function possessing a covariance function $R_g(v)$ defined by (1.6). *If*

$$R_g(v) \neq 0, \quad v = 0, 1, \ldots, \qquad \ldots \quad (1.14)$$

a consistent in quadratic mean estimate of the covariance function $R_Y(v)$ of the unobserved time series $Y(\bullet)$ is given by (1.11).

Assume next that the series $Y(\cdot)$ possesses a spectral density function $f_Y(\omega)$ so that

$$R_Y(v) = \int_{-\pi}^{\pi} \cos\, v\omega f_Y(\omega)\, d\omega. \qquad \ldots \quad (1.15)$$

Given consistent estimates $\hat{R}_Y(v)$ of $R_Y(v)$, one may construct consistent estimates $\hat{f}_Y(\omega)$ of $f_Y(\omega)$ in a multitude of ways by suitably choosing the weights $k_T(v)$ in the formula

$$\hat{f}_Y(\omega) = \frac{1}{2\pi} \hat{R}_Y(0) + \frac{1}{\pi} \sum_{v=1}^{T} \cos\, v\omega\, k_T(v)\, \hat{R}_Y(v); \qquad \ldots \quad (1.16)$$

proofs of this assertion are essentially given in Parzen (1961a) and Parthasarathy (1960). We do not discuss this assertion further here since we will actually obtain a formula for the asymptotic variance of the estimate $f_Y(\omega)$.

2. Missing observations

There exist time series $\{Y(t),\ t = 1, 2, \ldots\}$, defined at equally spaced intervals of time, which are systematically unobservable. For example, in radar studies of the surface of the moon, one observes a time series $Y(\cdot)$ which represents the *echo* (reflection from the moon) of a radar signal transmitted to the moon. In order to receive the echo, one must systematically cease transmission during the intervals in which one is receiving the echo. Another example of missing observations is the case of a time series which can be observed only during certain hours of the day.

A time series with missing observations seems to be best regarded as an amplitude modulated version of the original time series :

$$X(t) = g(t)\, Y(t), \quad t = 1, 2, \ldots, \qquad \ldots \quad (2.1)$$

where (i) $Y(\cdot)$ is the time series under study, assumed to be defined at successive equally spaced points of time, (ii) $g(\cdot)$ is defined by

$$\begin{aligned} g(t) &= 0 \text{ if } Y(t) \text{ is missing at time } t, \\ &= 1 \text{ if } Y(t) \text{ is observed at time } t, \end{aligned} \qquad \ldots \quad (2.2)$$

and (iii) $X(\cdot)$ represents the actually observed values of $Y(\cdot)$, with 0 inserted in the series whenever the value of $Y(t)$ is missing.

A case of particular interest is the case of systematically missing observations. Suppose that the time series $Y(\cdot)$ is periodically observed for α time points, then not observed for β time points; then $g(\cdot)$ is a periodic function with period $\alpha+\beta$, and

$$\begin{aligned} g(t) &= 1 \quad \text{if} \quad t = 1, 2, \ldots, \alpha, \\ &= 0 \quad \text{if} \quad t = \alpha+1, \ldots, \alpha+\beta. \end{aligned} \qquad \ldots \quad (2.3)$$

It may be shown that a periodic function $g(\bullet)$ possesses a generalized harmonic analysis. If the period of $g(\bullet)$ is θ (for $g(\bullet)$ defined by (2.3), $\theta=\alpha+\beta$), then its covariance function $R_g(\bullet)$ has period θ and is given by

$$R_g(v) = \frac{1}{\theta} \sum_{t=1}^{\theta} g(t)\, g(t+\theta). \qquad \ldots \quad (2.4)$$

Thus for $g(\bullet)$ defined by (2.3), the covariance function $R_g(\bullet)$ has period $\alpha+\beta$. To determine its values for $0 \leqslant v \leqslant \alpha+\beta-1$, we distinguish two cases : (i) $\alpha \leqslant \beta$ and (ii) $\alpha > \beta$.

TABLE 1. VALUES OF $R_g(v)$

case (i) : $\alpha \leqslant \beta$	case (ii) : $\alpha > \beta$
$\frac{\alpha-v}{\alpha+\beta}$, $v = 0, \ldots, \alpha,$	$\frac{\alpha-v}{\alpha+\beta}$, $v = 0, \ldots, \beta$
$0,\ v = \alpha, \ldots, \beta$	$\frac{\alpha-\beta}{\alpha+\beta}$, $v = \beta, \ldots, \alpha$
$\frac{v-\beta}{\alpha+\beta}$, $v = \beta, \ldots, \alpha+\beta$	$\frac{v-\beta}{\alpha+\beta}$, $v = \alpha, \ldots, \alpha+\beta.$

Only in the case $\alpha > \beta$ (one observes more values than one misses) does $R_g(v)$ never vanish. Therefore in order to be able to estimate $R_Y(v)$ we must assume that $\alpha > \beta$.

3. Variance of spectral estimates

In this section we find the variance of the estimated spectral density function $\hat{f}_Y(\omega)$ when it is formed from observations of an amplitude modulated time series $X(t)$ satisfying the assumptions of Theorem 1A. We first note that $\hat{f}_Y(\omega)$, defined by (1.16), can be written

$$\hat{f}_Y(\omega) = \frac{1}{2\pi T} \sum_{s,t=1}^{T} g(s)\, g(t)\, Y(s)\, Y(t)$$

$$\cos\, \omega(s-t)\ k_T(s-t)\ \{R_g(s-t)\}^{-1}. \qquad \ldots \quad (3.1)$$

In words, $\hat{f}_Y(\omega)$ is a quadratic form in the time series $Y(\bullet)$.

Let $a(s, t)$ be a symmetric function of two variables, and let

$$J_T|a(s, t)| = \sum_{s,t=1}^{T} a(s, t)\, Y(s)\, Y(t) \qquad \ldots \quad (3.2)$$

denote a quadratic form in the stationary normally distributed random variables $Y(\bullet)$ with covariance function $R_Y(\bullet)$ and spectral density function $f_Y(\bullet)$. It may be verified that

$$\operatorname{var}[J_T|a(s,t)|] = \sum_{s,t,u,v=1}^{T} a(s,t)\,a(u,v)\,\{R_Y(s-u)\,R_Y(t-v)+R_Y(s-v)\,R_Y(t-u)\}$$

$$= 2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} d\lambda_1\,d\lambda_2\,f_Y(\lambda_1)\,f_Y(\lambda_2)\,|A(\lambda_1,\lambda_2)|^2 \qquad \dots \quad (3.3)$$

defining

$$A(\lambda_1,\lambda_2) = \sum_{s,t=1}^{T} a(s,t)\,\exp[i(s\lambda_1+t\lambda_2)]. \qquad \dots \quad (3.4)$$

The usual case considered in the theory of spectral analysis of stationary time series is the case where

$$a(s,t) = \frac{1}{2\pi T}\cos\omega(s-t)\,k_T(s-t) \qquad \dots \quad (3.5)$$

and

$$k_T(v) = k(B_T v) \qquad \dots \quad (3.6)$$

for a suitable weighting function $k(v)$ and constants B_T satisfying

$$B_T \to 0, \quad TB_T \to \infty \text{ as } T \to \infty; \qquad \dots \quad (3.7)$$

for the exact conditions to be satisfied by the covariance averaging kernel $k(v)$ (see Parzen, 1957, p. 336). Define

$$K_T(\omega_1,\omega_2) = \frac{1}{2\pi T}\sum_{s,\,t=1}^{T} \exp[i(s\omega_1+t\omega_2)]k_T(s-t). \qquad \dots \quad (3.8)$$

By the argument employed in Parzen (1957, p. 342), one may show that for suitable functions $f(\lambda_1,\lambda_2)$ which are symmetric in the sense that

$$f(-\lambda_1,-\lambda_2) = f(\lambda_1,\lambda_2)$$

it holds that

$$\lim_{T\to\infty} TB_T \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(\lambda_1,\lambda_2)\;K_T(\lambda_1+\omega_1,\lambda_2+\omega_2)K_T(-\lambda_1-\omega_3,-\lambda_2-\omega_4)d\lambda_1 d\lambda_2$$

$$= \begin{cases} f(\omega_1,\omega_2)\displaystyle\int_{-\infty}^{\infty} k^2(u)du & \text{if } \omega_1=\omega_3,\ \omega_2=\omega_4. \\ 0 & \text{otherwise.}\end{cases} \qquad \dots \quad (3.9)$$

In particular, (3.9) holds for a function $f(\omega_1,\omega_2)$ of the form

$$f(\omega_1,\omega_2) = \frac{1}{2\pi}\sum_{v_1,\,v_2=-\infty}^{\infty} \exp[i(v_1\omega_1+v_2\omega_2)]\;R(v_1,v_2) \qquad \dots \quad (3.10)$$

where

$$\sum_{v_1,\,v_2=-\infty}^{\infty} |R(v_1,v_2)| < \infty. \qquad \dots \quad (3.11)$$

A careful derivation of (3.9) would unduly lengthen the present paper. However, let us sketch a proof. It suffices to show that (3.9) holds for

$$f(\lambda_1, \lambda_2) = \exp[i(\lambda_1 v_1+\lambda_2 v_2)] \qquad \dots \quad (3.12)$$

for arbitrary integers v_1 and v_2. Under (3.12), the double integral in (3.9) can be written

$$(4\pi^2 T)^{-1} B_T \sum_{s,\, t,\, u,\, v=1}^{T} k_T(s-t)\, k_T(u-v)\, J(s-u+v_1) J(t-v+v_2) \exp[i(s\omega_1+t\omega_2-u\omega_3-v\omega_4)], \qquad \dots \quad (3.13)$$

defining

$$J(\alpha) = \int_{-\pi}^{\pi} e^{i\alpha\lambda}\, d\lambda = 2\,\frac{\sin \pi\alpha}{\alpha}. \qquad \dots \quad (3.14)$$

In (3.13) make the change of variables

$$x = s-u, \quad y = t-v, \quad z = u-v.$$

so that

$$s = x+u, \quad t = y-z+u, \quad v = u-z.$$

Then (3.13) becomes

$$(4\pi^2)^{-1} B_T \sum_{x,\, y,\, z} J(x+v_1)\, J(y+v_2) k_T(z) k_T(z+x-y)$$

$$\exp[i\{x\omega_1+y\omega_2+z(\omega_4-\omega_2)\}] \times \frac{1}{T} \sum_{u=1}^{T} \exp\,[iu(\omega_1-\omega_3+\omega_2-\omega_4)]. \qquad \dots \quad (3.15)$$

As T tends to ∞, (3.15) has the following limiting values : if $\omega_1 = \omega_3$ and $\omega_2 = \omega_4$, then it has the value

$$(4\pi^2)^{-1} \sum_{x,\, y=-\infty}^{\infty} J(x+v_1)\, J(y+v_2) \exp[i(x\omega_1+y\omega_2)] \int_{-\infty}^{\infty} k^2(u) du \qquad \dots \quad (3.16)$$

and 0 otherwise. To conclude the proof of (3.9) we need only note that

$$(2\pi)^{-1} \sum_{x=-\infty}^{\infty} J(x+v_1) \exp[ix\omega_1] = \exp[-iv_1\omega_1). \qquad \dots \quad (3.17)$$

We next show how using (3.9) one may derive an expression for the variance of the spectral density function of an amplitude modulated normal time series. We are then considering quadratic forms corresponding to

$$a(s, t) = \frac{1}{2\pi T} \cos \omega(s-t)\, k_T(s-t)\, h(s, t), \qquad \dots \quad (3.18)$$

defining

$$h(s, t) = g(s)\, g(t)\, \{R_g(s-t)\}^{-1}. \qquad \dots \quad (3.19)$$

We consider only the important special case that $g(t)$ is a periodic function. If $g(t)$ has period θ, then it possesses the harmonic representation

$$g(t) = \sum_{n=-N}^{N} e_n \exp\,[it\, \lambda_n]\, G_n \qquad \dots \quad (3.20)$$

where $\lambda_n = 2\pi n/\theta$, $N = \theta/2$ or $(\theta-1)/2$ according as θ is even or odd, and

$$G_n = \frac{1}{\theta} \sum_{s=1}^{\theta} \exp[-is\lambda_n]\, g(s) \quad \text{for } n = 0, \pm 1, \dots, \pm[\theta/2], \qquad \dots \quad (3.21)$$

while $e_n = 1$ for all n except that if θ is even $e_{\pm N} = 1/2$. It may be verified that $R_g(v)$ is an even function of period θ given by

$$R_g(v) = \sum_{n=-N}^{N} e_n G_n G_{-n} \exp[-iv\lambda_n]. \qquad \dots \quad (3.22)$$

Now

$$g(s)\, g(t) = \sum_{m,\, n=-N}^{N} e_m e_n \exp\,[i(s\lambda_m + t\lambda_n)]\; G_m G_n$$

$$\{R_g(s-t)\}^{-1} = \sum_{m,\, n=-N}^{N} e_m e_n \exp\,[i(s\lambda_m + t\lambda_n)]\; W_{m,n} \qquad \dots \quad (3.23)$$

where, defining
$$W_n = \frac{1}{\theta} \sum_{s=1}^{\theta} \exp\,[-is\,\lambda_n]\{R_g(s)\}^{-1},$$

$$\begin{aligned} W_{m,n} &= W_n \quad \text{if } m = -n, \\ &= 0 \qquad \text{otherwise.} \end{aligned} \qquad \dots \quad (3.24)$$

Consequently
$$h(s,t) = \sum_{m,n=-N}^{N} e_m e_n \exp\,[i(s\lambda_m + t\lambda_n)] H_{m,n} \qquad \dots \quad (3.25)$$

where
$$H_{m,n} = \sum_{j,k} e_j\, e_k\, W_{j,k}\, G_{j-m}\, G_{k=n} = \sum_j e_j\, W_j\, G_{j-m}\, G_{-j-n}. \quad \dots \quad (3.26)$$

It should be noted that

$$\begin{aligned} H_{0,0} &= \Sigma\; e_j\, W_j\, G_j\, G_{-j} \\ &= \frac{1}{\theta} \sum_{s=1}^{\theta} R_g(s)\;\{(R_g(s)\}^{-1} \\ &= 1. \end{aligned} \qquad \dots \quad (3.27)$$

We next write

$$\begin{aligned} A(\lambda_1, \lambda_2) &= \frac{1}{2\pi T} \sum_{s,t} \cos \omega(s-t) k_T(s-t)\, h(s,t) \exp\,[i(s\lambda_1 + t\lambda_2)] \\ &= \frac{1}{2\pi T} \sum_{m,n=-N}^{N} e_m\, e_n\, H_{m,n} \sum_{s,t} \cos\; \omega(s-t)\; k_T(s-t) \\ &\qquad \exp\,[i(s\{\lambda_1+\lambda_m\} + t\{\lambda_2 + \lambda_n\})] \\ &= \sum_{m,n=-N}^{N} e_m\, e_n\, H_{m,n}\, \tfrac{1}{2}\{K_T(\lambda_1+\lambda_m+\omega,\, \lambda_2+\lambda_n-\omega) \\ &\qquad + K_T(\lambda_1+\lambda_m-\omega,\, \lambda_2+\lambda_n+\omega)\}. \end{aligned} \qquad \dots \quad (3.28)$$

We are now in a position to evaluate

$$TB_T \;\text{var}\;[\hat{f}_Y(\omega)] = 2TB_T \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} d\lambda_1\, d\lambda_2\, f_Y(\lambda_1) f_Y(\lambda_2)\, |A(\lambda_1, \lambda_2)|^2.$$

By (3.28) and (3.9) one sees that, as T tends to ∞,

$$TB_T \operatorname{var}[\hat{f}_Y(\omega)] \to \tfrac{1}{2} \sum_{m,n} e_m e_n |H_{m,n}|^2 \{f_Y(\lambda_m+\omega) f_Y(\lambda_n-\omega) + f_Y(\lambda_m-\omega) f_Y(\lambda_n+\omega)\} \int_{-\infty}^{\infty} k^2(u)du$$

$$= \left\{ \sum_{m,n} e_m e_n |H_{m,n}|^2 \, f_Y(\omega+\lambda_m) \, f_Y(\omega+\lambda_n) \right\} \int_{-\infty}^{\infty} k^2(u)du$$

$$= \left\{ f_Y^2(\omega) + \sum_{n \neq m} e_m e_n |H_{m,n}|^2 f_Y(\omega+\lambda_m) \, f_Y(\omega+\lambda_n \right\} \int_{-\infty}^{\infty} k^2(u)du. \quad \ldots \quad (3.29)$$

The foregoing formula is valid for $0 < \omega < \pi$; it should be multiplied by 2 in the case that $\omega = 0$ or $\omega = \pi$.

If the spectral density function $f_Y(\omega)$ had been directly estimated from observations of the time series $Y(\bullet)$, the variance of the estimate $\hat{f}_Y(\omega)$ would satisfy (for $0 < \omega < \pi$)

$$\lim_{T \to \infty} TB_T \operatorname{var}[\hat{f}_Y(\omega)] = f_Y^2(\omega) \int_{-\infty}^{\infty} k^2(u)du. \quad \ldots \quad (3.30)$$

Consequently one can infer from (3.29) the effect on the variance of the estimate $\hat{f}_Y(\omega)$ due to the fact that it is formed from an amplitude modulated version of the time series $Y(\bullet)$. An upper bound to this variance is

$$TB_T \operatorname{var}[\hat{f}_Y(\omega)] \leqslant \bar{H} \{\max_{\omega} f_Y^2(\omega)\} \int_{-\infty}^{\infty} k^2(u)du \quad \ldots \quad (3.31)$$

where

$$\bar{H} = \sum_{m,n} e_m e_n |H_{m,n}|^2. \quad \ldots \quad (3.32)$$

Thus $\bar{H}$ may be taken as a measure of the increase in variance due to amplitude modulation.

One may verify that

$$\bar{H} = \frac{1}{\theta^2} \sum_{s,t=1}^{\theta} h^2(s,t) = \frac{1}{\theta^2} \sum_{s,t=1}^{\theta} g^2(s)\, g^2(t) \, \{R_g(s-t)\}^{-2}. \quad \ldots \quad (3.33)$$

An upper bound for $\bar{H}$ can be obtained as follows. Let ρ be a lower bound for $R_g(v)$:

$$|R_g(v)| \geqslant \rho, \quad v = 0, 1, \ldots, \theta. \quad \ldots \quad (3.34)$$

Then

$$\bar{H} \leqslant \rho^{-2} \left\{ \frac{1}{\theta} \sum_{t=1}^{T} g^2(t) \right\}^2. \quad \ldots \quad (3.35)$$

An exact evaluation of $\bar{H}$ can be obtained from the formula

$$\bar{H} = \frac{1}{\theta} \sum_{v=-|\theta-1|}^{\theta-1} \{R_g(v)\}^{-2} \frac{1}{\theta} \sum_{t=1}^{\theta-|v|} g^2(t)\, g^2(t+|v|). \quad \ldots \quad (3.36)$$

To illustrate the use of these expressions we consider the modulating function $g(\circ)$, defined by (3.3), which corresponds to the case of periodically missing observations. Then $\theta = \alpha+\beta$,

$$\rho = \frac{\alpha-\beta}{\alpha+\beta}, \quad \frac{1}{\theta} \sum_{t=1}^{\theta} g^2(t) = \frac{\alpha}{\alpha+\beta}. \qquad \cdots \quad (3.37)$$

By (3.35)

$$\bar{H} \leqslant \left(\frac{\alpha}{\alpha-\beta}\right)^2 = \left(\frac{1}{1-r}\right)^2 \qquad \cdots \quad (3.38)$$

defining

$$r = \frac{\beta}{\alpha} \qquad \cdots \quad (3.39)$$

to be the ratio of the number of observations missed to the number observed. An exact expression for $\bar{H}$ can be obtained from (3.36) : for $v = 0, 1\ldots, \theta$

$$\begin{aligned} \frac{1}{\theta} \sum_{t=1}^{\theta-v} g^2(t)\, g^2(t+v) &= \frac{\alpha-v}{\alpha+\beta}, \quad v < \alpha, \\ &= 0, \qquad v \geqslant \alpha. \end{aligned} \qquad \cdots \quad (3.40)$$

Consequently,

$$\begin{aligned} \{R_g(v)\}^{-2} \frac{1}{\theta} \sum_{t=1}^{\theta-v} g^2(t)\; g^2(t+v) &= \frac{\alpha+\beta}{\alpha-v}, \quad v = 0, 1, \ldots, \beta; \\ &= \frac{(\alpha-v)(\alpha+\beta)}{(\alpha-\beta)^2}, \quad v = \beta, \ldots, \alpha; \\ &= 0, \quad v > \alpha. \end{aligned}$$

Thus
$$(\alpha+\beta)\,\bar{H} = \frac{\alpha+\beta}{\alpha} + 2 \sum_{v=1}^{\beta} \frac{\alpha+\beta}{\alpha-v} + 2 \sum_{v=\beta+1}^{\alpha} \frac{(\alpha-v)(\alpha+\beta)}{(\alpha-\beta)^2}.$$

Finally, one obtains

$$\begin{aligned} \bar{H} &= \frac{1}{\alpha} + \frac{\alpha-\beta-1}{\alpha-\beta} + 2\left\{\frac{1}{\alpha-1} + \frac{1}{\alpha-2} + \cdots + \frac{1}{\alpha-\beta}\right\} \\ &= \frac{1}{\alpha} + \frac{2}{\alpha-1} + \frac{2}{\alpha-2} + \cdots + \frac{2}{\alpha-\beta+1} + \frac{\alpha-\beta+1}{\alpha-\beta}. \end{aligned} \qquad \cdots \quad (3.41)$$

By replacing every denominator by $\alpha-\beta$, one obtains the following upper bound for $\bar{H}$:

$$\bar{H} \leqslant \frac{\alpha+\beta}{\alpha-\beta} = \frac{1+r}{1-r}. \qquad \cdots \quad (3.42)$$

One easily verifies that (3.42) provides a lower upper bound than does (3.38). In any event, both (3.38) and 3.42) provide some measure of how rapidly the variance of the spectral estimates increases as the ratio r tends to 1.

It may be of interest to express the variance of $\hat{f}_Y(\omega)$ in terms of the number T_0 of observations actually observed : approximately,

$$T_0 = \frac{\alpha}{\alpha+\beta} T. \qquad \ldots \quad (3.43)$$

Combining (3.31) and (3.42) one sees that

$$\frac{T_0 B_T \operatorname{var}[\hat{f}_Y(\omega)]}{\left\{\max\limits_{\omega} \hat{f}_Y(\omega)\right\}^2 \int\limits_{-\infty}^{\infty} k^2(u)du} \leqslant \frac{\alpha}{a+\beta} \overline{H} = \frac{1}{1-r}.$$

REFERENCES

DOOB, J. L. (1953) : *Stochastic Processes*, Wiley, New York.

JONES, R. H. (1962) : Spectral analysis with regularly missed observations. *Ann. Math. Stat.*, **32**, 455-461.

PARTHASARATHY, K. R. (1960) : On the estimation of the spectrum of a stationary stochastic process. *Ann. Math. Stat.*, **31**, 568-573.

PARZEN, E. (1957) : On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Stat.*, **28**, 329-348.

——— (1961a) : Mathematical considerations in the estimation of spectra *Technometrika*, **3**, 167-190.

——— (1961b) : Spectral analysis of asymptotically stationary time series. *Bulletin of the International Statistical Institute*, 33rd Session, Paris.

——— (1962) : *Stochastic Processes*, Holden-Day, San Francisco.

Paper received : October, 1962.

A METHOD OF CONSTRUCTION OF RESOLVABLE BIBD*

By ESTHER SEIDEN

Michigan State University

SUMMARY. A geometrical method of construction of resolvable BIBD with parameters $(v, b, r, k, \lambda) = (2^{2n-1}-2^{n-1}, 2^{2n}+1, 2^{n-1}, 1)$ is obtained, where n is a positive integer greater than or equal to two.

1. Introduction

A balanced incomplete block design, BIBD, with parameters v, b, r, k and λ is an arrangement of v treatments in b blocks of size k $(<v)$, such that no treatment occurs more than once in a block and every treatment occurs in exactly r blocks and every pair of treatments occurs together exactly λ times. The design is called resolvable (Bose, 1942) if the blocks can be divided into sets such that each treatment occurs in each set.

It is well known that a finite projective plane of order 2^n i.e. with 2^n+1 points on a line may contain a set of 2^n+2 points, three on one line called henceforth an oval. If such a plane is Desarguessian then ovals consiting of 2^n+2 points can be effectively constructed. This fact will be used in order to construct resolvable BIBD with parameters $(2^{2n-1}-2^{n-1},\ 2^{2n}-1,\ 2^n+1,\ 2^{n-1},\ 1)$ in which the $2^{2n}-1$ blocks can be divided into 2^n+1 sets each comprising 2^n-1 blocks such that each variety occurs in each of the sets.

2. Method of construction

It was shown by the author (Seiden, 1961) that the existence of an oval consisting of 2^n+2 points in a Desarguessian plane of order 2^n enables a construction of two partially balanced incomplete block designs, PBIBD, in the following manner. Remove the points of the oval. Then consider separately the lines which contained two points of the oval and the lines which contained none of them. If the lines are identified with the blocks of the design and the points with the varieties then one obtains this way two PBIBD.

Let us now focus our attention on the PBIBD consisting of the lines which did not contain any point of the oval. This design has $2^{2n-1}-2^{n-1}$ blocks of size 2^n+1, the number of varieties being $2^{2n}-1$. Every two blocks of this design will have clearly one element in common. It is easy to see that if we interchange now the roles of the points and the lines then we will obtain a BIBD consisting of $2^{2n}-1$ blocks. Each block corresponds to a point of the plane excluding the points of the oval. The

---

* This research memorandum was partially supported by the National Science Foundation Grant, No. G18376.

block size will be clearly 2^{n-1} and the elements of the block correspond to the lines not including any points of the oval which pass through the related point. It remains to establish that the so-obtained BIBD is in fact resolvable. Consider any point of the oval say O, and a line through it l. l contains 2^n-1 points not belonging to the oval. Each of these correspond to 2^n-1 different blocks of the BIBD. The 2^{n-1} elements of each block are the lines which do not include any point of the oval. Clearly they have to be all different and this will exhaust all $2^{2n-1}-2^{n-1}$ lines and give one replication of the design. The same will hold in respect to the remaining 2^n lines passing through O. Thus yielding the desired 2^n+1 sets of the resolvable design. Here is an example of a design obtained by the described method for $n = 3$.

1	2	3	4	1	5	6	7	1	8	9	10
5	15	22	24	2	8	20	24	2	12	18	22
6	9	12	21	3	12	23	28	3	16	25	27
7	16	18	28	4	10	16	21	4	15	20	28
8	14	23	26	9	11	15	19	5	19	21	26
10	13	19	25	13	14	18	27	6	13	17	23'
11	17	20	27	17	22	25	26	7	11	14	24

1	11	12	13	1	14	15	16	1	17	18	19
2	7	9	25	2	10	11	26	2	6	15	27
3	6	20	26	3	13	21	24	3	5	9	14
4	19	24	27	4	7	12	17	4	8	13	22
5	8	16	17	5	18	20	25	7	10	20	23
10	14	22	28	6	8	19	28	11	21	25	28
15	18	21	23	9	22	23	27	12	16	24	26

1	20	21	22	1	23	24	25	1	26	27	28
2	16	19	23	2	5	13	28	2	14	17	21
3	8	11	18	3	10	15	17	3	7	19	22
4	6	14	25	4	9	18	26	4	5	11	33
5	10	12	27	6	11	16	22	6	10	18	24
7	13	15	26	7	8	21	27	8	12	15	25
9	17	24	28	12	14	19	20	9	13	16	20

After the paper was submitted for publication R. C. Bose informed me that he was aware of this result since 1949 and used it in a proof in his joint paper with S. S. Shrikhande. [R. C. Bose and Shrikhande (1960) : On the construction of sets of mutually orthogonal latin squares and the falsity of a conjecture of Euler, *Transaction of the American Mathematical Society*, **95**, 191-209.]

REFERENCES

BOSE, R. C. (1942) : A note on resolvability of balanced incomplete block designs. *Sankhyā*, **6**, 105-110.

SEIDEN, ESTHER (1961) : On a geometrical method of construction of partially balanced designs with two associate classes. *Ann. Math. Stat.*, **32**, 1117-1180.

Paper received : April, 1963.

SOME LIMIT DISTRIBUTIONS CONNECTED WITH FIXED INTERVAL ANALYSIS*

By J. SETHURAMAN

Indian Statistical Institute

and

University of North Carolina

SUMMARY. The proofs of some theorems stated by the author (Sethuraman, 1963) on the limiting distributions of some statistics that enter in the method of Fixed Interval Analysis are presented.

1. Introduction

Let $(\boldsymbol{Y}, X)$ be a random variable taking values in $(\mathscr{Y} \times \mathscr{X})$ where $\mathscr{Y}$ is E_k the Euclidean space of k dimensions and $\mathscr{X}$ is a measurable space. Let $E_1, E_2, \ldots, E_g$ be g disjoint measurable sets in $\mathscr{X}$ whose union is the whole space $\mathscr{X}$.

$(\boldsymbol{y}_1, x_1), (\boldsymbol{y}_2, x_2), \ldots, (\boldsymbol{y}_n, x_n)$ are n independent observations on $(\boldsymbol{Y}, X)$. The number of x_i's that fall in E_j is n_j, $j = 1, \ldots, g$. $\boldsymbol{u}_j$ is defined by the relation

$$\boldsymbol{u}_j = \Sigma^j \boldsymbol{y}_i / n_j \qquad j = 1, \ldots, g$$

where Σ^j is the summation over all "i" such that x_i is in E_j.

Throughout this paper it is assumed that

$$V(\boldsymbol{Y}) < \infty \qquad \ldots \quad (1.1)$$

and

$$\text{prob } (X \epsilon E_j) = \pi_j > 0 \qquad j = 1, \ldots, g \qquad \ldots \quad (1.2)$$

where for any random variable $\boldsymbol{Z}$, $v(\boldsymbol{Z})$ denotes the variance covariance matrix of $\boldsymbol{Z}$.

The following theorem is established in Section 3.

Theorem 1 : *The asymptotic distribution of* $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_g)$ *is the distribution of* g *independent normal distributions.*

This theorem plays a fundamental role in the method of Fixed Interval Analysis (for instance, see Sethuraman (1963)). Interpreted in Sample Survey language this theorem, among other things, states that the post-stratified stratum means are independently distributed in the limit.

2. Notations, definitions and preliminaries

Let $\boldsymbol{Y}(E_j)$, called the conditional random variable of $\boldsymbol{Y}$ given that X is in E_j, denote a random variable on $\mathscr{Y}$ with the distribution defined by prob $(\boldsymbol{Y}(E_j) \epsilon A) = \text{prob } (\boldsymbol{Y} \epsilon A, \; X \epsilon E_j)/\text{prob } (X \epsilon E_j)$. For any random variable $\boldsymbol{Z}$, $E(\boldsymbol{Z})$ denotes the vector of expectations of $\boldsymbol{Z}$.

---

* This research work was done at the Indian Statistical Institute. The writing of this paper was supported by the Mathematics Division of the Air Force Office of Scientific Research under grant number AF-FOSR-62-169.

Define

$$E(\boldsymbol{Y}(E_j)) = \boldsymbol{\mu}_j \qquad \dots \quad (2.1)$$

$$V(\boldsymbol{Y}(E_j)) = \Sigma_j \qquad \dots \quad (2.2)$$

$$p_j = n_j/n \qquad \dots \quad (2.3)$$

$$\sqrt{n}(\boldsymbol{u}_j - \boldsymbol{\mu}_j) = \boldsymbol{\eta}_j(n) \qquad \dots \quad (2.4)$$

$$\sqrt{n}(\boldsymbol{p}_j - \boldsymbol{\pi}_j) = \boldsymbol{\zeta}_j(n) \qquad \dots \quad (2.5)$$

$$j = 1, \dots, g.$$

Let $\{\xi_n(\cdot, \theta)\}$, $n = 0, 1, \dots$ be a sequence of families of probability distributions on the Borel-subsets of E_m (or more generally, of any topological space) and θ vary in a compact topological space K.

Definition: $\{\xi_n(\cdot, \theta)\}$ is said to converge weakly, uniformly and continuously (in other words, in the UC^* sense) to $\xi_0(\cdot, \theta)$ with respect to θ in K if for every bounded continuous function $h(y)$ on E_m

$$\int g(y)\ \xi_n(dy, \theta) \rightarrow \int g(y)\ \xi_0(dy, \theta) \text{ uniformly in } \theta$$

and

$$\int g(y)\ \xi_0(dy, \theta) \text{ is a continuous function of } \theta.$$

The following theorem given by the author (Sethuraman, 1961) will be used in Section 3.

Theorem 2 : *Let (Y_n, X_n) be a sequence of random variable on $(E_m \times S)$ where S is a complete separable metric space. Let the conditional probability measure of Y_n given that $X_n = x$ be denoted by $\xi_n(\cdot, x)$ and the marginal distribution of X_n be μ_n. Let $\xi_n(\cdot, x)$ converge in the UC^* sense to $\xi_0(\cdot, x)$ with respect to x in any compact subset of S and μ_n converge weakly to μ_0. Then the joint distribution of (Y_n, X_n) converges weakly to the distribution determined by $\xi_0(\cdot, x)$ and μ_0 or, more precisely, to the distribution of (Y_0, X_0) where*

$$\text{prob}\{Y_0 \in A,\quad X_0 \in B\} = \int_B \xi_0(A, x)\mu(dx).$$

Lemma which is immediate, is useful in establishing the UC^* convergence of a special sequence of families of distributions.

Let

$$\boldsymbol{Z}_{11}, \dots, \boldsymbol{Z}_{1k_1(\theta)}$$

$$\boldsymbol{Z}_{21}, \dots, \boldsymbol{Z}_{2k_2(\theta)}$$

$$\dots\dots\dots\dots\dots\dots$$

$$\boldsymbol{Z}_{n1}, \dots, \boldsymbol{Z}_{nk_n(\theta)}$$

be a triangular scheme of random variables in E_m where the variables in any row are identically and independently distributed. Assume that $E(\boldsymbol{Z}_{n1}) = \boldsymbol{\nu}_n$ and $V(\boldsymbol{Z}_{n1}) = V_n$ are finite and that $V_n \rightarrow V$ as $n \rightarrow \infty$. Again let $\inf_\theta k_n(\theta) \rightarrow \infty$ as $n \rightarrow \infty$.

Let $MN(\boldsymbol{\alpha}, L)$ stand for the multivariate normal distribution with mean vector $\boldsymbol{\alpha}$ and variance covariance matrix L.

Lemma 1 : *The sequence of families of distributions of*

$$\{(\boldsymbol{Z}_{n_1}+\ldots+\boldsymbol{Z}_{nk_n(\theta)}-k_n(\theta)\boldsymbol{\nu}_n)/\sqrt{k_n(\theta)}\}$$

converges in the UC sense to the distribution $MN(0, V)$ with respect to θ.*

3. Main theorems

We first prove the following lemma.

Lemma 2 : *The distributions of $(\eta_1(n), \ldots, \eta_g(n))$ given that $\zeta(n) = z$, $\sum_1^g z_i = 0$ converges in the UC* sense to the distribution $MN(0, \Lambda)$ with respect to z in any closed bounded subset of E_g, where*

$$\Lambda = \begin{pmatrix} \frac{1}{\pi_1}\Sigma_1 & 0 & 0 \\ 0 & \frac{1}{\pi_2}\Sigma_2 & 0 \\ \cdots & \cdots & \cdots \\ 0 & 0 & \frac{1}{\pi_g}\Sigma_g \end{pmatrix} \qquad \ldots \quad (3.1)$$

Proof: The event $\zeta(n) = z$ is equivalent with probability one to the event $n_i = [n\pi_i + \sqrt{n}\, z_i]$ $i = 1, \ldots, g$, since

$$\text{prob}\ \{n\pi_i + \sqrt{n}\zeta_i(n) = [n\pi_i + \sqrt{n}\zeta_i(n)],\quad i = 1, \ldots, g\} = 1.$$

The conditional distribution of $y_1, \ldots, y_n$ given that $n_i = [n\pi_i + \sqrt{n}\, z_i]$, $i = 1, \ldots, g$ is the distribution of g independent samples of size $n_1, \ldots, n_g$ on $\boldsymbol{Y}(E_1), \ldots, \boldsymbol{Y}(E_g)$, respectively. $\sqrt{\frac{n_1}{n}}\,\eta_1(n), \ldots, \sqrt{\frac{n_g}{n}}\,\eta_g(n)$ are the normalized means of these g independent samples. For z in a closed bounded subset of E_g we note that $\inf_{i, z} [n\pi_i + \sqrt{n}\, z_i] \to \infty$ as $n \to \infty$. Thus all the conditions of Lemma 1 are satisfied. Further $[n\pi_i + \sqrt{n}\, z_i)/n$ tends to π_i, uniformly in z in any closed bounded subset of E_g. Hence the conditional distributions of $(\eta_1(n), \ldots, \eta_g(n))$ given that $\zeta(n) = z$ converges in the UC^* sense to the distribution $MN(0, \Lambda)$ with respect to z in any closed bounded subset of E_g.

Theorem 3 : *The joint distribution of $(\eta_1(n), \ldots, \eta_g(n), \zeta(n))$ converges weakly to the distribution $MN(0, B)$*

where
$$B = \begin{pmatrix} \Lambda & 0 \\ 0 & C \end{pmatrix} \qquad \ldots \quad (3.2)$$

$$\text{and where} \quad C = \begin{vmatrix} \pi_1(1-\pi_2) & -\pi_1\pi_2 & \dots & -\pi_1\pi_g \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & \dots & -\pi_2\pi_g \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot & \cdot \\ -\pi_1\pi_g & -\pi_2\pi_g & \dots & \pi_g(1-\pi_g) \end{vmatrix} \quad \dots \quad (3.3)$$

Proof: This theorem is an immediate consequence of Theorem 2, Lemma 2 and the observation that the distribution of $\zeta(n)$ converges weakly to the distribution $MN(0, C)$.

Proof of Theorem 1 : Theorem 1 is contained in Theorem 3.

REFERENCES

SETHURAMAN, J. (1961) : Some limit theorems for joint distributions. *Sankhyā*, Series A, **23**, 379–386.

———(1963) : Fixed interval analysis and fractile analysis. *Contributions to Statistics*, Presented to Professor P. C. Mahalanobis on the occasion of his 70th birthday, Pergamon Press, Oxford and Statistical Publishing Society, Calcutta.

Paper received : February, 1963.

A METHOD OF CONSTRUCTION OF INCOMPLETE BLOCK DESIGNS

By S. S. SHRIKHANDE

Banaras Hindu University

and

D. RAGHAVARAO*

Bombay University

SUMMARY. A new method of constructing α-resolvable and affine α-resolvable balanced incomplete block (BIB) and partially balanced incomplete block (PBIB) designs is given. Two series of BIB designs with $b = 4(r-\lambda)$ are also obtained.

1. Introduction

An incomplete block design with parameters v, b, r and k is an arrangement of v treatments in b blocks of size $k(< v)$, such that no treatment occurs more than once in a block and further every treatment occurs in exactly r blocks. An incomplete block is said to be balanced if every pair of treatments occurs in exactly λ blocks. If we can define an association scheme for the treatments as given by Bose and Mesner (1959) and if the pair of treatments which are i-th associates occur together in λ_i blocks ($i = 1, 2, \ldots, m$), then the design is called a PBIB design if all the λ's are not equal. These designs were first introduced by Yates (1936-1937) and Bose and Nair (1939) respectively.

An incomplete block design with parameters v, b, r and k is said to be α-resolvable (Shrikhande and Raghavarao, 1962) if the blocks can be divided into t sets $S_1, S_2, \ldots, S_t$ each of β blocks, such that in each set every treatment is replicated α times. We then necessarily have

$$v\alpha = k\beta, \quad r = \alpha t, \quad b = \beta t. \qquad \ldots \quad (1.1)$$

An α-resolvable incomplete block design will be said to be affine α-resolvable, if every pair of blocks of the same set intersect in q_1 treatments whereas any pair of blocks from different sets intersect in q_2 treatments. A 1-resolvable and an affine 1-resolvable design may be simply called resolvable and affine resolvable design respectively.

We have shown that an α-resolvable or an affine α-resolvable design can be constructed, given two equireplicate designs D_1 and D_2 where D_2 is resolvable or affine resolvable and has the same number of blocks in each replication as the number of treatments in D_1. This method of construction is presented in the next section.

2. Method of construction

Let D_1 be an incomplete block design with parameters v_1, b_1, r_1, k_1 and D_2 be a resolvable incomplete block design with parameters $v_2 = k_2v_1$, $b_2 = r_2v_1$, r_2, k_2.

* The work of this author was financially supported by the Government of India Research Fellowship.

Let the blocks of the j-th replicate S_j of D_2 be arbitrarily numbered $1, 2, \ldots, v_1$ ($j = 1, 2, \ldots, r_2$). Let $M = (m_{ij})$ be the usual incidence matrix of D_1 where $m_{ij} = 1$ or 0 according as the i-th treatment of D_1 occurs or does not occur in the j-th block of D_1. Let M_j be the incidence matrix obtained from M by replacing each treatment i of D_1 by the set of treatments contained in the i-th block of S_j. It is then easy to see that

$$N = (M_1\ M_2, \ldots, M_{r_2}) \qquad \ldots \quad (2.1)$$

is the incidence matrix of an α-resolvable incomplete block design D with parameters

$$v = v_2, \quad b = b_1 r_2, \ r = r_1 r_2, \quad k = k_1\, k_2 \qquad \ldots \quad (2.2)$$
$$\alpha = r_1, \beta = b_1, \quad t = r_2,$$

where each set of blocks M_j is an α-replicate of D.

We now consider some particular cases of the above construction.

Case 1 : *D_1 and D_2 BIB designs.* Let D_1 be a BIB design with parameters $v_1, b_1, r_1, k_1, \lambda_1$ and D_2 be a BIB design with parameters $v_2 = k_2 v_1$, $b_2 = r_2 v_1$, r_2, k_2, λ_2. Then as shown above we get a design D with parameters given by (2.2). Further, it is easy to verify that any two treatments of D occur together in exactly $r_1\lambda_2 + \lambda_1(r_2 - \lambda_2)$ blocks. Thus D is a BIB design. Hence we have

Theorem 1 : *The existence of a* BIB *design D_1 with parameters*

$$v_1, b_1, r_1, k_1, \lambda_1 \qquad \ldots \quad (2.3)$$

and a resolvable BIB *design with parameters*

$$v_2 = k_2 v_1, \ b_2 = r_2 v_1, \ r_2, \ k_2, \ \lambda_2 \qquad \ldots \quad (2.4)$$

implies the existence of an α-resolvable BIB *design D with parameters*

$$v = v_2, \ b = b_1 r_2, \ r = r_1 r_2, \ k = k_1 k_2$$
$$\lambda = r_1\lambda_2 + \lambda_1(r_2 - \lambda_2), \ \alpha = r_1, \ \beta = b_1, \ t = r_2 \qquad \ldots \quad (2.5)$$

In particular, consider the case when D_1 is a symmetrical BIB design and D_2 is an affine resolvable BIB design. We note that any two blocks of D_1 intersect in λ_1 treatments and any two blocks of different replications in D_2 intersect in $\frac{k_2^2}{v_2}$ treatments (Bose, 1942). It is then easy to verify that in D any two blocks of the set M_j intersect in $\lambda_1 k_2$ treatments and any two blocks from different sets M_i and M_j intersect in $\frac{k_1^2 k_2}{v_1}$ treatments. Thus D is an affine α-resolvable BIB design. Since the parameters of an affine resolvable BIB design can be expressed in terms of two parameters $n \geqslant 2$, $m \geqslant 0$ (Bose, 1942), we have the following :

Theorem 2 : *The existence of a symmetrical* BIB *design with parameters*

$$v_1 = b_1 = n, \qquad r_1 = k_1, \lambda_1 \qquad \ldots \quad (2.6)$$

and an affine resolvable BIB *design with parameters*

$$v_2 = nk_2 = n^2[(n-1)m+1], \ b_2 = nr_2 = n(n^2m+n+1), \ \lambda_2 = nm+1 \quad \ldots \quad (2.7)$$

implies the existence of an affine α-resolvable BIB *design with parameters*

$$v = v_2,\ b = b_2,\ r = r_1r_2,\ k = k_1k_2,\ \lambda = r_1\lambda_2+\lambda_1(r_2-\lambda_2)$$
$$\alpha = r_1,\ \beta = n,\ t = r_2. \qquad \ldots \quad (2.8)$$

If s and $p = s^2+s+1$ are both prime powers, then we can take D_1 to be PG(2, s) amd D_2 to be EG(2, p). Hence we get the following :

Corollary : *If s and $p = s^2+s+1$ are both prime powers, then we can construct an affine α-resolvable* BIB *design with parameters*

$$v = p^2,\ b = p(p+1),\ r = (p+1)(s+1),\ k = p(s+1),\ \lambda = s+p+1$$
$$\alpha = s+1,\ \beta = p,\ t = p+1. \qquad \ldots \quad (2.9)$$

Case 2 : *D_1 a* BIB *design and D_2 a* PBIB *design.* Let D_1 be a BIB design with parameters $v_1, b_1, r_1, k_1, \lambda_1$ and D_2 be a resolvable PBIB design with m associate classes having the association scheme A and with parameters

$$v_2 = k_2v_1, \quad b = r_2v_1,\ r_2,\ k_2,\ \lambda_{2,1},\ \ldots\ \lambda_{2,m}. \qquad \ldots \quad (2.10)$$

Then analogous to Theorems 1 and 2, we can prove the following theorem :

Theorem 3 : *The existence of a* BIB *design D_1 with parameters* (2.3) *and a resolvable* PBIB *design D_2 with parameters* (2.10) *implies the existence of an α-resolvable* PBIB *design with the same association scheme* A *and with parameters*

$$v^* = v_2, \quad b^* = r_2b_1,\ r^* = r_1r_2,\ k^* = k_1\,k_2,$$
$$\lambda_i^* = r_1\lambda_{2,i}+\lambda_1(r_2-\lambda_{2,i}), \quad i = 1, 2, \ldots m \qquad \ldots \quad (2.11)$$
$$\alpha = r_1,\ \beta = b_1,\ t = r_2.$$

If in particular D_1 is symmetrical and D_2 is affine resolvable, then D is affine α-resolvable.

3. BIB designs of the family (A) with $b = 4(r-\lambda)$

Shrikhande (1962) has shown that BIB designs of the family (A) with $b=4(r-\lambda)$ have an interesting reproducing property in that any two members of this family give rise to another member of the same family. In this section we give two series belonging to the family.

When $4m+3$ is a prime power then we know that the following designs exist (Bose, 1939).

$$D_1 : v_1 = b_1 = 4m+3,\ r_1 = k_1 = 2m+1,\ \lambda_1 = m$$
$$D_2 : v_2 = v_1^2,\ b_2 = v_1^2+v_1,\ r_2 = v_1+1,\ k_2 = v_1,\ \lambda_2 = 1.$$

Using Theorem 2 we get

$$D : \quad v = v_1^2,\ b = v_1^2+v_1,\ r = \frac{v_1^2-1}{2},\ k = \frac{v_1(v_1-1)}{2}$$

$$\lambda = \frac{(v_1+1)(v_1-2)}{4} \qquad \ldots \quad (3.1)$$

which obviously belongs to the family (A).

Similarly if $4m+1$ is a prime power, then the following designs exist :

$$D_1 : v_1 = 4m+1,\ b_1 = 2v_1,\ r_1 = 4m,\ k_1 = 2m,\quad \lambda_1 = 2m-1$$

$$D_2 : v_2 = v_1^2,\quad b_2 = v_1^2+v_1,\quad r_2 = v_1+1,\ k_2 = v_1,\ \lambda_2 = 1$$

Hence Theorem 1 gives

$$D : v = v_1^2,\ b = 2v_1(v_1+1),\quad r = v_1^2-1,\ k = \frac{v_1(v_1-1)}{2} \qquad \dots \quad (3.2)$$

$$\lambda = \frac{(v_1+1)(v_1-2)}{2}$$

which again belongs to the family (A). It is interesting to note that (3.2) can also be obtained by the method of block intersection (Bose, 1939) from the symmetrical BIB design with parameters

$$v = b = v_1^2+(v_1+1)^2,\ r = k = v_1^2,\ \lambda = \frac{v_1(v_1-1)}{2}$$

REFERENCES

BOSE, R. C. (1939) : On the construction of balanced incomplete block designs. *Ann. Eugen.*, **9**, 353–399.

——— (1942) : A note on the resolvability of balanced incomplete block designs. *Sankhyā*, **6**, 105–110.

BOSE, R. C. and MESNER, DALE, M. (1959) : On linear associative algebras corresponding to association schemes of partially balanced designs. *Ann. Math. Stat.*, **30**, 21–38.

BOSE, R. C. and NAIR, K. R. (1939) : Partially balanced incomplete block designs. *Sankhyā*, **4**, 337–372.

SHRIKHANDE, S. S. (1962) : On a two parameter family of balanced incomplete block designs. *Sankhyā*, Series A, **24**, 33–40.

SHRIKHANDE, S. S. and RAGHAVARAO, D. (1964) : Affine α-resolvable incomplete block designs. *Contributions to Statistics*, Volume presented to Professor P. C. Mahalanobis on his 70th birthday; Pergamon Press, Oxford and Statistical Publishing Society, Calcutta.

YATES, F. (1936-37) : Incomplete randomized blocks. *Ann. Eugen.* **7**, 121–140.

Paper received : March, 1963.

FIDUCIAL DISTRIBUTIONS ASSOCIATED WITH INDEPENDENT NORMAL VARIATES

By D. A. SPROTT

University of Waterloo

SUMMARY. Several different fiducial distributions that have occurred in the literature, some of which are controversial, are closely related to the distribution of independent normal variates and owe their existence to a property of this distribution. This note brings these fiducial distributions together comparing their derivations and contrasting some of their properties.

1. Introduction

There have been several examples in the past of non-uniqueness of fiducial distributions. These arise from different methods of factoring the distribution of the sample often, but not always, by means of a change of variable. Some of these different fiducial distributions are characterized uniquely by a set of transformations that will produce them (Fraser, 1961a, b) indicating that such transformations may form an essential logical aspect of the specification and information along with the likelihood function. For example, the bivariate normal distribution gives rise to two "regression" distributions, characterized by two different groups of transformations corresponding to x independent, y dependent variable and conversely; it also gives rise to a symmetric fiducial distribution (Fisher, 1956; Fraser, 1961a, b; Mauldon, 1955; Quenouille, 1958; Sprott 1961).

Brillinger (1962) cites two examples on non-uniqueness rising from two different ways of factoring the distribution of the sufficient statistics without any change of variable involved, that is

$$\begin{aligned} f(S_1, S_2; \theta_1, \theta_2) &= f(S_2; \theta_2) f(S_1; \theta_1, \theta_2 \mid S_2) \\ &= f(S_1; \theta_1) f(S_2; \theta_1, \theta_2 \mid S_1). \end{aligned}$$

An example of this for three parameters is provided again by the bivariate normal distribution, which can be written

$$\begin{aligned} f(S_1, S_2, r; \sigma_1, \sigma_2, \rho) &= f(S_1, S_2; \sigma_1, \sigma_2, \rho \mid r) f(r; \rho) \\ &= f(S_1; \sigma_1) f(r; \sigma_1, \rho \mid S_1) f(S_1; \sigma_1, \sigma_2, \rho \mid r, S_1) \end{aligned}$$

It is the purpose of this note to bring together and discuss some different and apparently unrelated fiducial distributions, some of which are controversial, that owe their existence to the fact that this property is possessed by the simple example of two independent normal variates x, y with means μ, ν and variances unity. The parameters of particular concern are the ratio $\alpha = \mu/\nu$ and the distance $\rho = \sqrt{\mu^2+\nu^2}$. Unlike the previous examples however, some of the results are not proper distributions. It will be found that one set of distributions, characterized by location transformations, are obtained by conditioning on y; another set, characterized in part by rotation transformations are obtained by conditioning on $r = \sqrt{x^2+y^2}$.

2. INDEPENDENT NORMAL VARIATES : LOCATION TRANSFORMATIONS

Here $$f(x, y; \mu, \nu) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}(x-\mu)^2 - \frac{1}{2}(y-\nu)^2\right];$$

setting $a = x/y$, $\alpha = \mu/\nu$, this can be written

$$\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(y-\nu)^2\right] \frac{1}{\sqrt{2\pi}}|y| \exp\left[-\frac{1}{2}(ay-\alpha\nu)^2\right] = f(y; \nu) f(a; \alpha, \nu \mid y). \quad \ldots \quad (1)$$

The first factor leads to the usual fiducial normal distribution of ν, the second to the distribution of α given ν as normal with mean $\frac{x}{\nu}$ and variance unity. Thus the simultaneous distribution of α, ν is

$$\frac{1}{2\pi}|\nu| \exp\left[-\frac{1}{2}(ay-\alpha\nu)^2 - \frac{1}{2}(y-\nu)^2\right], \quad \ldots \quad (2)$$

and integrating with respect to ν produces the distribution of α given by Creasy (1954) about which there is some controversy.

A similar argument (conditioning on y) produces the distribution of ρ,

$$\frac{1}{2\pi}\rho \exp\left[-\frac{1}{2}(r^2+\rho^2)\right] \int_0^{2\pi} \exp(r\rho \cos u)du, \quad \ldots \quad (3)$$

about which there is some controversy (Stein, 1959; James, 1954). Both of these distributions are related to the usual fiducial distribution of μ, ν and are usually derived by integrating it.

All of the above distributions can be obtained empirically by sampling the normal distribution (x', y') and applying location transformations $x' \to x'+c_1 = x$, $y' \to y'+c_2 = y$, $\mu \to \mu+c_1 = \mu'$, $\nu \to \nu+c_2 = \nu'$, e.g. Fraser (1961 a, b) giving

$$\frac{\mu}{\nu} = \alpha \to \frac{\mu+(x-x')}{\nu+(y-y')} = \frac{\mu'}{\nu'} = \alpha', \rho \to \sqrt{\mu'^2+\nu'^2} = \rho'$$

where μ', ν' are independent normal variates with means x, y and variances unity.

3. INDEPENDENT NORMAL VARIATES : ROTATION TRANSFORMATIONS

The distribution (1) can be integrated with respect to y by setting $y = r \sin t$, $\nu = \rho \sin \theta$, $a = \cot t$, and $\alpha = \cot \theta$. This gives

$$f(r; \rho) = \frac{1}{2\pi} r \exp\left[-\frac{1}{2}(r^2+\rho^2)\right] \int_0^{2\pi} \exp(r\rho \cos u)du \quad \ldots \quad (4)$$

and $$f(t; \rho, \nu \mid r) = \exp[r\rho \cos(t-\theta)] / \int_0^{2\pi} \exp(r\rho \cos u)\, du. \quad \ldots \quad (5)$$

Thus the parent distribution can be conditioned on r

$$f(a, y; \alpha, \nu) = f(r, t; \rho, \theta) = f(r; \rho)\, f(t; \rho, \theta \mid r). \quad \ldots \quad (6)$$

The first factor leads to a distribution (not proper) of ρ (James, 1954 ; Stein, 1959), and the second to the distribution of θ given ρ,

$$f(\theta; t, r \mid \rho) = \exp r\rho \cos(\theta-t) / \int_0^{2\pi} \exp(r\rho \cos u)\, du. \quad \ldots \quad (7)$$

This in fact produces the distribution of μ, ν conditional on the circle $\mu^2+\nu^2 = \rho^2$ (Fisher, 1956) and thus the distribution of α given ρ.

The product of (7) with $f(\rho) = -\frac{\partial}{\partial\rho} F(r, \rho)$ derived from (4) leads to a simultaneous distribution of θ,ρ which does not produce the usual normal distribution of μ, ν (although combining (7) with $f(\rho)$ given by (3) does). Thus the distribution of θ,ρ produced this way is in conflict with the distribution (2) of α, ν. The conditional distribution of $\alpha = \cot\theta$ can be found from (7) as

$$f(\alpha\,|\,\rho) = 2\cosh r\rho \frac{1+a\alpha}{\sqrt{1+a^2}\,\sqrt{1+\alpha^2}} \Big/ (1+\alpha^2) \int_0^{2\pi} \exp\,(r\rho\cos u)\,du \quad \ldots \quad (8)$$

Multiplying this by $f(\rho)$ above and integrating would give a new marginal distribution of α different from before. Multiplying $f(\alpha\,|\,\rho)$ above by the expression for $f(\rho)$ derived previously (3) and integrating reproduces the Creasy (1954) distribution for α as in the preceding section.

The conditional distributions $f(\alpha\,|\,\rho)$, $f(\nu\,|\,\rho)$ in this section can be obtained empirically by sampling the normal distribution conditional on a given circle and applying rotation transformations. That is, given the original pair of observations $(x, y) = (r, t)$ $(x^2+y^2 = r^2)$ the normal distribution could be sampled for a new pair $(x', y') = (r, t')$ subject to $x'^2+y'^2 = r^2$. The rotation transformation $t' \to t = t'+(t-t')$ carries t' into t and $\theta = \cot^{-1}\mu/\nu$ into $\theta' = \theta+(t-t')$. Thus $\theta' = \cot\alpha'$ is a random variable with distribution given by (7), so that α' is a random variable with conditional distribution given by (8). It does not appear that the marginal distribution of α can be given a frequency interpretation this way, as no such transformations seem applicable to r to produce the distribution of ρ derived from (4).

4. Discussion

It can be seen from the foregoing that the distribution of r, y can be factored in two ways (conditional on y, and conditional on r) giving rise to two sets of possible fiducial distributions of ρ, α, μ, ν etc. One of these sets is characterized by location transformations and can be entirely reproduced by them; the other is characterized by rotation transformations and can be reproduced partly (conditional on ρ) by them.

The question arises as to which, of these sets of distribution has more claim to validity. Those produced in Section 2 by location transformations have many desirable consistency properties. They reproduce the usual normal fiducial distribution of μ, ν and thus serve to preserve the symmetry with respect to the actual experiment whereby an interchange of x with μ, y with ν transforms the distribution of x, y into the fiducial distribution of μ, ν. Thus it would seem reasonable that the logical status of x, y before the experiment for a known μ, ν is identical to that of μ, ν after the experiment for a known x, y. Then any statements appropriate to x, y, $x/y = a$ and $\sqrt{x^2+y^2} = r$ etc. before the experiment should correspond completely to similar statements about μ, ν, α ρ etc., after, and should be derivable by the above interchange. In this respect the particular observation $x = y = 0$ would cause no more difficulty in making probability statements about α than would the knowledge that $\mu = \nu = 0$ in making statements about a (both giving rise to the Cauchy distribution). Also, all the distributions in Section 2 are proper distributions and can be

produced empirically by sampling the normal distribution and applying location transformations. None of these points is true for the distributions arising in Section 3 although for a fixed ρ the conditional distributions can be produced by rotation transformations. The distribution of ρ in Section 3 seems to be considered more valid by some (James, 1954; Stein, 1959); some of the associated conditional distributions have occurred before in the literature in a different context (e.g. distributions on a circle, Fisher, 1956).

It is interesting to note that the commonly-accepted distribution of α based on the standard normal variate (Fieller, 1954).

$$u = (x-\alpha y)/\sqrt{1+\alpha^2} \qquad \ldots \quad (9)$$

apparently has none of the above properties, nor does it seem to be associated as closely with the parent normal distribution by the usual fiducial argument presented by Fisher (1956). In spite of the fact it is invariant under the group of rotations, it does not appear to be derivable from it in the manner of Fraser (1961a, b) and Sections 2, 3. For if x', y' are sampled from a normal distribution, they cannot necessarily be transformed into the original pair (x, y), the group of rotations not being transitive. For this to be so, (x', y') must be sampled conditionally on the circle $x'^2+y'^2 = r^2 = x^2+y^2$. Then u is no longer a standard normal variate so that the result of applying these transformations is apparently given by (8) and not (9). It is curious that the distributions $f(\rho)$ in Section 3 and $f(\alpha)$ (9) for which some preference is shown in the literature do not exhibit any of the consistency features of the distributions derived in Section 2 which, seem desirable from the point of view of transformations and mathematical manipulation of ordinary probability distributions.

It is possible that the distribution $f(\rho) = \dfrac{-\partial F(r, \rho)}{\partial \rho}$ obtained from (6) can be derived empirically by a set (though not a group) of transformations in the manner described by Sprott (1963). In any particular problem it would then be necessary to decide which group or set of transformations were relevant, as pointed out in Section 1. For example, the hypothesis $\rho = 0$ (or $\rho = \rho_0$) is in conflict with the location transformations that produce $f(\nu)$ in Section 2, for then ν^2 must be zero (or less than ρ_0^2). Thus location transformations and the resulting distribution (3) do not seem appropriate for such a test (cf. Barnard, 1963).

REFERENCES

BARNARD, G. A., (1963) : Logical aspects of the fiducial argument. 34th Session of *Int. Stat. Inst.*, Ottawa.

BRILLINGER, D. R. (1962) : Examples bearing on the definition of fiducial probability with a biblography. *Ann. Math.* Stat., **33**, 1349-1355.

CREASY, M. A. (1954) : Symposium on interval estimation. *J. Roy. Stat., Soc.*, B, **16**, 186-194.

FIELLER, E. C. (1954) : Symposium on interval estimation. *J. Roy. Stat. Soc.*, B. **16**, 175-185.

FISHER, R. A. (1956) : *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinbourgh and London.

FRASER, D. A. S. (1961a) : On fiducial inference. *Ann. Math. Stat.*, **32**, 616-676.

FRASER, D. A. S. (1961b) : The fiducial method and invariance. *Biometrika*, **48**, 261-280.

JAMES, G. S. (1954) : Discussion on symposium on interval estimation *J. Roy. Stat., Soc.*, B, **16**, 214-216.

MAULDON, J. G. (1955) : Pivotal quantities for Wishart's and related distributions and a paradox in fiducial theory. *J. Roy. Stat., Soc.*, B, **17**, 79-85.

QUENOUILLE, M. H. (1958) : *The Fundamentals of Statistical Reasoning*, Charles Griffin, London.

SPROTT, D. A. (1961) : Similarities between likelihoods and associated distributions a priori. *J. Roy. Stat., Soc.*, B, **23**, 460-468.

——— (1963) : A transformation model for the investigation of fiducial distributions. 34th Session of the *Int. Stat, Inst.* Ottawa.

STEIN, C. (1959) : An example of wide discrepancy between fiducial and confidence intervals. *Ann. Math. Stat.*, **30**, 877-880.

Paper received : February, 1963.

MULTIVARIATE STATISTICAL OUTLIERS*

By S. S. WILKS

Princeton University

SUMMARY. This paper deals with the problem of identifying and testing a candidate set of a small number t of extreme sample elements as significant outliers in a sample of size n from a k-dimensional normal distribution with unknown parameters. The problem is considered in detail for $t=1, 2, 3, 4$, that is, for sets of 1, 2, 3, and 4 outliers. The criterion for indentifying and testing a single observation as a significant outlier is r_1 as defined in Section 3(b) and that for a pair of outliers is r_2 as defined in Section 4, small values of r_1 or r_2 being critical values. In the absence of exact values for the extremely complicated probabilities $P(r_1 < r)$ and $P(r_2 < r)$ upper bounds for these probabilities are given by (2.15) and (3.2) respectively. These upper bounds are suggested for *a fortiori* significance testing of observed values of r_1 and r_2. Some evidence of the closeness of these upper bounds obtained for the probabilities $P(r_1 < r)$ and $P(r_2 < r)$ is given in Table 1 for $k=1$, that is, for a sample from a one-dimensional normal distribution. In this case exact values of r_α for which $P(r_1 < r_\alpha) = \alpha$ are available from Grubbs' (1950) tables for certain values of α. These are compared with the upper bounds of $P(r_1 < r_\alpha)$ for several values of n in Table 1.

Values of r_α for which the upper bound of $P(r_1 < r_\alpha)$ has the value α are given in Table 2 for $\alpha = 0.010, 0.025, 0.050, 0.100$; $k=1, 2, 3, 4, 5$; and $n=5(1)30(5)100(100)500$. Table 3 gives values of $\sqrt{r_\alpha}$ for which the upper bound of $P(r_2 < r_\alpha)$ has the value α for the same values of α, k and n.

Extension of r_1 and r_2 to the case of t outliers is r_t as defined in Section 5. Expressions are given for the cases $t=3$ and 4 from which values of r_α can be determined so that the upper bound of $P(r_t < r_\alpha)$ is α. No tabulations have been made, however, for the cases of three and four outliers.

In the more general problem of t outliers a procedure is outlined as to how one could obtain the value of r_α for which the upper bound of $P(r_t < r_\alpha)$ has the value α.

1. Introduction

Studies of criteria for the rejection of extreme observations as significant outliers in a single sample from a one-dimensional normal distribution with unknown parameters have been made by various authors during the last thirty years.

If $(x_1, \ldots, x_n)$ is a sample from such a distribution and if $\bar{x}$ and s^2 are the sample mean and sample variance, Thompson (1935) has determined the distribution of $(x_\xi - \bar{x})/s$ for an arbitrary ξ. He has proposed that for a given α, values of x_ξ for which $|x_\xi - \bar{x}|/s > \tau_\alpha$ be rejected as significant outliers in a sample from a normal distribution where τ_α is chosen so that for any ξ, $P(|x_\xi - \bar{x}|/s > \tau_\alpha) = \alpha$. He determined τ_α for $\alpha = \frac{0.05}{n}, \frac{0.10}{n}, \frac{0.20}{n}$ and for $n = 3(1)22, 32, 42, 102, 202, 1002$. Thus, for instance, if $\alpha = \frac{0.10}{n}$ the expected number of observations which would be falsely rejected as outliers, (that is, would be rejected if all elements of the sample were actually from the same normal distribution) would be 1 per 10 samples of size n.

* Research partially supported by the Office of Naval Research while the author was a Fellow of the Center for Advanced Study in the Behavioral Sciences in the Fall of 1961. Presented at the International Congress of Mathematicians, Stockholm, August 20, 1962.

Pearson and Chandra Sekar (1936) considered $(x_{(n)}-\bar{x})/s$ and $(\bar{x}-x_{(1)})/s$ as criteria for rejecting individual observations as significantly high and low outliers respectively, where $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$ are the order statistics of the sample. In particular, they showed that the upper tail of the distribution of $(x_{(n)}-\bar{x})/s$ (or of $(\bar{x}-x_{(1)})/s$) has a density function $xf_n(\tau)$ on the interval $(\sqrt{(n-2)/2},\ \sqrt{n-1})$, where $f_n(\tau)$ is the probability density function of $(x_\xi-\bar{x})/s = \tau$, say. From this fact they found the upper 1%, 2.5% and 10% points of the distribution of $(x_{(n)}-\bar{x})/s$ (or of $(\bar{x}-x_{(1)})/s$) for values of n ranging from 11 to 19, that is, for all values of n such that the specified upper percentage point falls in the interval $(\sqrt{(n-2)/2},\ \sqrt{n-1})$.

Grubbs (1950) extended the work of Pearson and Chandra Sekar (1936) for individual outliers by actually determining the distribution of $(x_{(n)}-\bar{x})/s$ (or of $(\bar{x}-x_{(1)})/s$) in a sample from a normal distribution with unknown parameters. He tabulated the upper 1%, 2.5%, 5% and 10% points of the distribution of $(x_{(n)}-\bar{x})/s$ (or of $(\bar{x}-x_{(1)})/s$) for all $n \leqslant 25$. He also tabulated the lower 1 %, 2.5%, 5% and 10% points of the distribution of $\sum_{\xi=1}^{n-1}(x_{(\xi)}-\bar{x}_n)^2/[(n-1)s^2]$ (or of $\sum_{\xi=2}^{n}(x_{(\xi)}-\bar{x}_1)^2/[(n-1)s^2]$) for all $n \leqslant 25$ where $\bar{x}_n$ is the mean of $x_{(1)}, \ldots, x_{(n-1)}$ and $\bar{x}_1$ is the mean of $x_{(2)}, \ldots, x_{(n)}$. Grubbs also considered the case of two high (or two low) outliers, using as the criterion of rejection $\sum_{\xi=1}^{n-2}(x_{(\xi)}-\bar{x}_{n,\ n-1})/[(n-2)s^2]$ (or $\sum_{\xi=3}^{n}(x_{(\xi)}-\bar{x}_{1,2})/[(n-2)s^2]$) where $\bar{x}_{n,\ n-1}$, is the mean of $x_{(1)}, \ldots, x_{(n-2)}$ and $\bar{x}_{1,2}$ is the mean of $x_{(\xi)}, \ldots, x_{(n)}$. He tabulated the lower 1%, 2.5%, 5% and 10% points of the distribution of these quantities for all $n \leqslant 20$. He mentioned, but did not go into the details of $\sum_{\xi=2}^{n-1}(x_{(\xi)}-\bar{x}_{1,n})^2/[(n-2)s^2]$ as a two-outlier test, where $\bar{x}_{1,n}$ is the mean of $x_{(2)}, \ldots, x_{(n-1)}$.

Dixon (1951) has considered ratios of form $(x_{(n)}-x_{(n-j)})/(x_{(n)}-x_{(i)})$ [or $(x_{(j+1)}-x_{(1)})/(x_{(n-i+1)}-x_{(1)})$], $i = 1, 2, 3$; $j = 1, 2$, as criteria for testing extreme observations as outliers and he has tabulated the 0.5%, 1%, 2%, 5%, 10(10)90%, 95% points of the distributions of these quantities. Dixon (1950) has also studied the power functions of all of the criteria mentioned above against alternatives in which it is assumed that the outliers are from normal distributions of form $N(\mu+\lambda\sigma, \sigma^2)$ or $N(\mu, \lambda^2\sigma^2)$ for various values of λ and for unknown μ and σ^2.

All of the studies mentioned above deal with the problem of testing one or two extreme observations as significant outliers in a sample from a one-dimensional normal distribution with unknown parameters.

Problems of outliers in samples from normal distributions for which one or both of the parameters are known or are estimated from independent samples have been considered by various authors, including Irwin (1925), McKay (1935), Newman (1940), Pearson and Hartley (1942), Nair (1948, 1952), David (1956), Pillai and Tienzo (1959) and Pillai (1959). Rider (1932) has given a survey of the literature on outliers prior to 1932.

The purpose of the present paper is to discuss in detail and present tables for the problem of selecting and testing one or two extreme observations as significant outliers in a sample from a multivariate normal distribution, with unknown parameters. The mathematical theory of selecting and testing three or more extreme observations as significant outliers is discussed, but no tables are given.

No attempt has been made to study the power of the outlier tests discussed in this paper under various possible alternatives to the null hypothesis that all of the elements of the sample are independently drawn from a common k-dimensional normal distribution with unknown parameters. This would be a much more extensive investigations than the study of the tests presented in this paper under the null hypothesis. Such a study remains to be done. Some of the power properties of a test equivalent to r_1 the test for the problem of one-outlier, have been investigated by Karlin and Truax (1960), and by Ferguson (1961).

2. The case of a single outlier

(a) *The one-outlier scatter ratios of a sample.* Let $(x_{1\xi}, \ldots, x_{k\xi}\ ;\ \xi = 1, \ldots, n)$ be a sample of size n from a k-dimensional normal distribution $N(\{\mu_i\}, \|\sigma_{ij}\|)$ where $\{\mu_i\}$ is the vector of means $(\mu_1, \ldots, \mu_k)$ and $\|\sigma_{ij}\|$ is the covariance matrix of the distribution. It is assumed that the vector of means and covariance matrix of the distribution are unknown. Let $(\bar{x}_1, \ldots, \bar{x}_k)$ be the vector of sample means, where $n\bar{x}_i = \sum_{\xi=1}^{n} x_{i\xi}$ and let

$$a_{ij} = \sum_{\xi=1}^{n} (x_{i\xi} - \bar{x}_i)(x_{j\xi} - \bar{x}_j), \qquad i, j = 1, \ldots, k. \qquad \ldots \quad (2.1)$$

The sample can be represented as a cluster of n points in a k-dimensional euclidean space R_k. Any k of these n points together with the sample center of gravity point $(\bar{x}_1, \ldots, \bar{x}_k)$ forms a simplex. If the volume of this simplex is squared and if the sum of squares is taken of the volumes of all possible simplexes which can be formed in this manner, it can be shown (see Wilks, 1962, for instance) that this sum of squared volumes is

$$(k!)^{-2}|a_{ij}|, \qquad \ldots \quad (2.2)$$

where $|a_{ij}|$ is the determinant of the matrix $\| a_{ij} \|$. It is convenient to call $|a_{ij}|$ the *internal scatter* of the sample $(x_{1\xi}, \ldots, x_{k\xi}\ ;\ \xi = 1, \ldots, n)$; if $n > k$, $| a_{ij} | > 0$ with probability 1.

If we delete the ξ-th element of the sample we obtain a cluster of $n-1$ points in R_k. Let the internal scatter of these $n-1$ points be $|a_{ij\xi}|$ which will be > 0 with probability 1 if $n > k+1$.

Let

$$R_\xi = \frac{|a_{ij\xi}|}{|a_{ij}|}, \quad \xi = 1, \ldots, n. \qquad \ldots \quad (2.3)$$

The quantities $R_1, \ldots, R_n$ will be called *one-outlier scatter ratios* of the sample $(x_{1\xi}, \ldots, x_{k\xi}\ ;\ \xi = 1, \ldots, n)$.

It can be verified that

$$a_{ij\xi} = a_{ij} - b_{i\xi} b_{j\xi} \qquad \dots \quad (2.4)$$

where

$$b_{i\xi} = \sqrt{\frac{n}{n-1}}\,(x_{i\xi} - \bar{x}_i).$$

Thus we have

$$|a_{ij\xi}| = |a_{ij} - b_{i\xi}\, b_{\;\xi}| = |a_{ij}| . [1 - \sum_{i,j=1}^{k} a^{ij} b_{i\xi} b_{j\xi}], \qquad \dots \quad (2.5)$$

where

$$\| a^{ij} \| = \| a_{ij} \|^{-1}.$$

Hence

$$R_\xi = 1 - \sum_{i,j=1}^{k} a^{ij} b_{i\xi} b_{j\xi}$$

and since

$$\sum_{\xi=1}^{n} a^{ij} b_{i\xi} b_{j\xi} = \frac{nk}{(n-1)},$$

we have

$$\sum_{\xi=1}^{n} R_\xi = n\left(1 - \frac{k}{n-1}\right). \qquad \dots \quad (2.6)$$

Now, it is known in multivariate statistical analysis (see Wilks (1962), for example) that for any ξ the ratio R_ξ has the beta distribution $B_e\left(\frac{n-k-1}{2}, \frac{k}{2}\right)$, where a random variable z is said to have the beta distribution $B_e(\nu_1, \nu_2)$ if the probability density function of z is

$$f(z) = \frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)}\, z^{\nu_1-1}\, (1-z)^{\nu_2-1} \qquad \dots \quad (2.7)$$

on the interval $(0, 1)$ and $f(z) = 0$ outside the interval.

Under the null hypothesis (that is, assuming that all elements of the sample are independently drawn from a common k-dimensional normal distribution), the one-outlier scatter ratios $R_1, \dots, R_n$ are random variables having a distribution which is symmetric over the n-dimensional space of $R_1, \dots, R_n$ for which

$$R_1 + \dots + R_n = n\left(1 - \frac{k}{n-1}\right) \qquad \dots \quad (2.8)$$

$$0 \leqslant R_\xi \leqslant 1, \qquad \xi = 1, \dots, n,$$

where the (marginal) distribution of each R_ξ is identical with the distribution of a random variable u having the beta distribution $B_e\left(\frac{n-k-1}{2}, \frac{k}{2}\right)$.

(b) *The ordered values of one-outlier scatter ratios.* Let $R_{(1)} < \dots < R_{(n)}$ be the ordered values of $R_1, \dots, R_n$. The criterion we propose for selecting and testing a single extreme observation as a significant outlier is $R_{(1)}$ which we shall denote by r_1. In other words the strongest candidate for being a significant outlier is identified as the sample element whose deletion gives the scatter ratio $r_1 = \min_\xi \{R_\xi\}$. It is the one to be tested as a significant outlier, with r_1 being the test criterion. It is evident that the critical values of r_1 are those in the left tail of its distribution.

The joint distribution of $R_{(1)}, \ldots, R_{(n)}$, or even of $R_1, \ldots, R_n$ for that matter, is very complicated. However, one can readily obtain moments of any one of the random variables $R_1, \ldots, R_n$ and also certain low joint moments of two or more of these random variables. For instance,

$$\mathcal{E}(R_\xi) = \frac{n-k-1}{n-1}, \ \mathrm{var}\,(R_\xi) = \frac{2k(n-k+1)}{(n-1)^2(n+1)}$$

$$\mathrm{cov}\,(R_\xi, R_\eta) = -\frac{2k(n-k+1)}{(n-1)^3(n+1)}. \qquad \cdots \quad (2.9)$$

Even though it does not appear feasible to determine exact percentage points in the lower tail of the distribution of r_1, except for $k=1$ and then only for small values of n as we shall see later, we can determine upper bounds for the amount of probability in the lower tail of the distribution of r_1 which should be useful, at least for small values of k and small percentage points, for *a fortiori* significance testing of r_1.

First let us examine the lower limits of the ranges of $R_{(1)}, \ldots, R_{(n)}$. If we consider the space of $(R_1, \ldots, R_n)$ remembering that $R_1, \ldots, R_n$ must each lie on the interval $(0, 1)$ it will be seen that not more than $n-k-1$ of the R's, in the set $\{R_1, \ldots, R_n\}$ can be 1 simultaneously. For if this were possible the average of the remaining R's in this set would be negative. This means that $R_{(1)}$ would be negative contrary to the fact that each R in the set $\{R_1, \ldots, R_n\}$ must lie on the interval $(0, 1)$. Thus if $n-k-1$ R's in the set $\{R_1, \ldots, R_n\}$ are simultaneously equal to 1, we would have $R_{(1)}+\cdots+R_{(k+1)} = k+1-\dfrac{nk}{n-1}$ and hence the average of $R_{(1)}, \ldots, R_{(k+1)}$ would be $1-\dfrac{nk}{(k+1)(n-1)}$ which implies that $R_{(k+1)} \geqslant 1-\dfrac{nk}{(k+1)(n-1)}$. Similarly, if we put $n-k-2$ of the R's in the set $\{R_1, \ldots, R_n\}$ equal to 1, it will be seen that $R_{(k+2)} \geqslant 1-\dfrac{nk}{(k+2)(n-1)}$. Continuing this process, if we put only one R in the set $\{R_1, \ldots, R_n\}$ equal to 1, we obtain

$$R_{(n)} \geqslant 1-\frac{k}{n-1}.$$

Note that it is possible for $R_{(1)}, \ldots, R_{(k)}$ to be 0 simultaneously, in which case $R_{(k+1)} = 1-\dfrac{nk}{(k+1)(n-1)}$. Therefore for left-hand end points of the distribution of $R_{(1)}, \ldots, R_{(n)}$ we have :

$$\begin{aligned} R_{(1)} &\geqslant 0, \ldots, R_{(k)} \geqslant 0 \\ R_{(k+1)} &\geqslant 1-\frac{nk}{(k+1)(n-1)} \\ R_{(k+2)} &\geqslant 1-\frac{nk}{(k+2)(n-1)} \\ &\vdots \\ R_{(n)} &\geqslant 1-\frac{k}{n-1}. \end{aligned} \qquad \cdots \quad (2.10)$$

(c) *Upper bound for* $P(r_1 < r)$. For a fixed number r let us consider the problem of finding an upper bound for $P(r_1 < r)$. Let $E_1, \ldots, E_n$ denote the events for which $R_1 < r, \ldots, R_n < r$ respectively. Then

$$P(r_1 < r) = P(E_1 \cup \cdots \cup E_n). \qquad \ldots \quad (2.11)$$

But
$$P(E_1 \cup \cdots \cup E_n) \leqslant P(E_1) + \cdots + P(E_n), \qquad \ldots \quad (2.12)$$

and
$$P(E_1) = \cdots = P(E_n) = P(u < r), \qquad \ldots \quad (2.13)$$

where
$$P(u < r) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-k-1}{2}\right)\Gamma\left(\frac{k}{2}\right)} \int_0^r u^{\frac{n-k-1}{2}-1} (1-u)^{\frac{k}{2}-1} du, \qquad \ldots \quad (2.14)$$

since, as stated in Section 2(a), u is a random variable having the beta distribution $B_e\left(\frac{n-k-1}{2}, \frac{k}{2}\right)$.

Therefore
$$P(r_1 < r) \leqslant nP(u < r), \qquad \ldots \quad (2.15)$$

that is, *$nP(u < r)$ is an upper bound for $P(r_1 < r)$*. In particular, if we choose $r = r_\alpha$ so that $nP(u < r_\alpha) = \alpha$ we obtain

$$P(r_1 < r_\alpha) \leqslant \alpha. \qquad \ldots \quad (2.16)$$

(d) *The upper bound $nP(u < r)$ as the expected number of scatter ratios with values $< r$.* The quantity $nP(u < r)$ has another useful interpretation. Suppose δ_ξ is a random variable which has the value 1 if $R_\xi < r$ and 0 otherwise, $\xi = 1, \ldots, n$. Let

$$N(r) = \sum_{\xi=1}^{r} \delta_\xi, \qquad \ldots \quad (2.17)$$

that is, $N(r)$ is the number of the one-outlier scatter ratios which have values less than r. We have

$$\mathcal{E}(N(r)) = \mathcal{E}(\delta_1) + \ldots + \mathcal{E}(\delta_n) = nP(u < r) \qquad \ldots \quad (2.18)$$

since $\mathcal{E}(\delta_\xi) = P(u < r)$, $\xi = 1, \ldots, n$. Thus, *the expected number of the one-outlier scatter ratios $R_1, \ldots, R_n$ having values less than r is equal to the upper bound $nP(u < r)$ of $P(r_1 < r)$.*

In particular, we have

$$\mathcal{E}(N(r_\alpha)) = P(r_1 < r_\alpha) = \alpha. \qquad \ldots \quad (2.19)$$

(e) *Comparison of values of upper bound of $P(r_1 < r)$ with Grubbs' exact values of $P(r_1 < r)$ for a sample from a one-dimensional normal distribution.* For the case $k = 1$ it will be seen from (3.10) that $R_{(1)} \geqslant 0$ (i.e. $r_1 \geqslant 0$), and $R_{(2)} \geqslant 1 - \frac{n}{2(n-1)}$. Hence for any value of r on the interval $\left(0, 1-\frac{n}{2(n-1)}\right)$ the expression (2.15) is an equality. For a value of r which exceeds $1-\frac{n}{2(n-1)}$ expression (2.15) is a strict inequality. In this case r_1 is the smaller of the two quantities $\sum_{\xi=1}^{n-1} (x_{(\xi)} - \bar{x}_n)^2/[(n-1)s^2]$ and $\sum_{\xi=2}^{n} (x_{(\xi)} - \bar{x}_1)^2/[(n-1)s^2]$ which were considered by Grubbs (1950) as criteria for upper

and lower outliers in a sample from a one-dimensional distribution. Thus, if r_α is the lower $100\alpha\%$ point of r_1 and lies on the interval $\left(0, 1-\frac{n}{2(n-1)}\right)$ it is the lower $100\frac{\alpha}{2}\%$ of each of the two criteria considered by Grubbs. For the case $k = 1$ Table 1 gives a comparison between the probability $P(r_1 < r_\alpha)$ and its upper bound $nP(u < r_\alpha)$ for $\alpha = 0.02$, 0.05, 0.10 and 0.20 and for certain values of n from Grubbs' tables for which inequality (2.15) is a strict inequality.

TABLE 1. COMPARISON OF $P(r_1 < r_\alpha)$ WITH ITS UPPER BOUND $nP(u < r_\alpha)$

n	α	r_α	exact probability (by Grubbs) $P(r_1 < r_\alpha)$	upper bound $nP(u < r_\alpha)$ of $P(r_1 < r_\alpha)$ [or equivalently, $E(N(r_\alpha))$]
20	.02	.5393	.020	.020
25		.6071	.020	.021
15		.5030	.050	.050
20	.05	.5937	.050	.050
25		.6544	.050	.052
15		.5558	.100	.100
20	.10	.6379	.100	.100
25		.6922	.100	.103
10		.4881	.200	.200
15	.20	.6134	.200	.200
20		.6848	.200	.206
25		.7319	.200	.210

(f) *Tables of values of r_α for which upper bound $nP(u < r_\alpha) = \alpha$.* For the case $k \geqslant 2$ it will be seen from (3.10) that the left hand endpoints of the distributions of $R_{(1)}, \ldots, R_{(k)}$ are all 0. Therefore for $k \geqslant 2$ expression (2.15) is a strict inequality; and there exists no value of r for which $nP(u < r)$ provides an exact value of $P(r_1 < r)$. The problem of determining exact values of $P(r_1 < r)$ for $k \geqslant 2$ does not seem feasible at present because of the complexity of the distribution of r_1. We therefore resort to the use of the upper bound $nP(u < r)$.

Table 2 gives values of r_α for which the upper bound $nP(u < r_\alpha)$ of $P(r_1 < r_\alpha)$ has the value α[or equivalently, values of r_α for which $\mathcal{E}(N(r_\alpha)) = \alpha$] for $\alpha = 0.010$, 0.025, 0.050, 0.100; $k = 1, 2, 3, 4, 5$; and $n = 5(1)30(5)100(100)500$.

3. The case of two outliers

Suppose we delete two elements, say $(x_{1\xi}, \ldots, x_{k\xi})$ and $(x_{1\eta}, \ldots, x_{k\eta})$ from the sample defined in Section 2 and denote the internal scatter of the resulting cluster of $n-2$ points by $|a_{ij\xi\eta}|$ which is positive with probability 1 if $n > k+2$. Let

$$R_{\xi\eta} = \frac{|a_{ij\xi\eta}|}{|a_{ij}|}, \ \eta > \xi = 1, \ldots, n. \qquad \ldots \quad (3.1)$$

The quantities $\{R_{\xi\eta}\}$ will be called *two-outlier scatter ratios* of the sample $(x_{1\xi}, \ldots x_{k\xi};$ $\xi = 1, \ldots, n)$. The conditions satisfied by the $\{R_{\xi\eta}\}$ except that each must lie on (0, 1) appear rather complicated and no attempt will be made here to state them.

It can be shown (see Wilks (1962), for instance) that for $n > k+2$ each of the $\binom{n}{2}$ scatter ratios in the set $\{R_{\xi\eta}\}$ has the property that its distribution is identical with that of a random variable u^2 where u has the beta distribution $B_e(n-k-2, k)$.

Let $r_2 = \min\limits_{\eta > \xi} \{R_{\xi\eta}\}$. The criterion proposed here for selecting and testing the strongest candidate pair of sample elements as significant outliers is r_2, that is, the candidate pair whose deletion in computing two-outlier scatter ratios produces the smallest scatter ratio.

No attempt is made here to give inequalities for these ordered scatter ratios analogous to those for the $\{R_{(1)}, \ldots, R_{(n)}\}$ as given in (2.10).

Under the null hypothesis, (that is, assuming that all elements in the sample are independently drawn from a common k-dimensional normal distribution) the joint distribution of $\{R_{\xi\eta}, \eta > \xi = 1, \ldots, n\}$ is symmetric in the $R_{\xi\eta}$, although apparently very complicated. However, an upper bound for the probability $P(r_2 < r)$ can be found by a procedure similar to that by which (2.15) was established, namely

$$P(r_2 < r) \leqslant \binom{n}{2} P(u^2 < r) \qquad \ldots \quad (3.2)$$

where
$$P(u^2 < r) = \frac{\Gamma(n-2)}{\Gamma(n-k-2)\,\Gamma(k)} \int_0^r (\sqrt{u})^{n-k-3}(1-\sqrt{u})^{k-1} d\sqrt{u}, \qquad \ldots \quad (3.3)$$

remembering that each $R_{\xi\eta}$ is a random variable having a distribution identical to that of a random variable u^2 where u has the beta distribution $B_e\,(n-k-2,\ k)$.

In particular if we choose r_α such that

$$\binom{n}{2} P(u^2 < r_\alpha) = \alpha, \qquad \ldots \quad (3.4)$$

we have
$$P(r_2 < r_\alpha) \leqslant \alpha. \qquad \ldots \quad (3.5)$$

As in the one-outlier problem, if we let $N(r)$ be the number of the $\binom{n}{2}$ two-outlier scatter ratios $\{R_{\xi\eta}\}$ which have values less than r, then

$$\mathcal{E}(N(r)) = \binom{n}{2} P(u^2 < r). \qquad \ldots \quad (3.6)$$

In particular, we have

$$\mathcal{E}(N(r_\alpha)) = \binom{n}{2} P(u^2 < r_\alpha) = \alpha. \qquad \ldots \quad (3.7)$$

Values of $\sqrt{r_\alpha}$ for which the upper bound $\binom{n}{2} P(u^2 < r_\alpha)$ of $P(r_2 < r_\alpha)$ has the value α [or equivalently, values of $\sqrt{r_\alpha}$ for which $\mathcal{E}[N(r_\alpha)) = \alpha$] are given in Table 3 for $\alpha = 0.010,\ 0.025,\ 0.050,\ 0.100$; $k = 1, 2, 3, 4, 5$; and $n = 5(1)\,30\,(5)\,100(100)500$.

4. The case of three or more outliers

The scatter ratio criteria for selecting and testing outliers can be extended to the case of three or more outliers in a fairly straightforward way.

For t outliers, we define the $\binom{n}{t}$ t-outlier scatter ratios as

$$R_{\xi_1 \dots \xi_t} = \frac{|a_{ij\xi_1 \cdots \xi_t}|}{|a_{ij}|}, \qquad \dots \quad (4.1)$$

$\xi_t < \dots < \xi_1 = 1, \dots, n$ where $|a_{ij\xi_1 \cdots \xi_n}|$ is the internal scatter of the $n-t$ points remaining in the sample after deletion of $(x_{1\xi_1} \dots, x_{k\xi_1}), \dots, (x_{1\xi_t} ; \dots , x_{\mu k\xi_t})$. The scatter ratio $R_{\xi_1 \dots \xi_t}$ is positive with probability 1 if $n > k+t$. The smallest of these scatter ratios, which we denote by r_t is the proposed criterion for selecting the t most extreme observations in the sample and for testing this set of t observations as a set of significant outliers.

Under the assumption that the n elements in the sample are independently drawn from a common k-dimensional normal distribution any one of the scatter ratios, say $R_{\xi_1 \dots \xi_t}$ is a random variable whose k-th moment is given by (see Wilks (1962))

$$\mathcal{E}(R^h_{\xi_1 \cdots \xi_t}) = \prod_{i=1}^{t} \frac{\Gamma\left(\frac{n-i}{2}\right)\Gamma\left(\frac{n-k-i}{2}+h\right)}{\Gamma\left(\frac{n-i}{2}+h\right)\Gamma\left(\frac{n-k-i}{2}\right)}, \qquad \dots \quad (4.2)$$

$$h = 0, 1, 2, \dots$$

Note that the h-th moment of $R_{\xi_1 \cdots \xi_t}$ is identical with the h-th moment of the product $z_1 \dots z_t$ where $z_1, \dots, z_t$ are independent random variables having beta distributions $B_e\left(\frac{n-k-1}{2}, \frac{k}{2}\right), \dots, B_e\left(\frac{n-k-t}{2}, \frac{k}{2}\right)$, respectively. The distribution of $R_{\xi_1 \dots \xi_t}$, is uniquely determined by its moments (see Cramér (1943)). Hence the distribution of $R_{\xi_1 \dots \xi_t}$ is identical with the distribution of the product $z_1 \dots z_t$ and hence

$$P(R_{\xi_1 \cdots \xi_t} < r) = P(z_1 \dots z_t < r). \qquad \dots \quad (4.3)$$

As in the one- and two-outlier problems we find that

$$P(r_t < r) \leqslant \binom{n}{t} P(z_1 \dots z_t < r), \qquad \dots \quad (4.4)$$

the probability $P(z_1 \dots z_t < r)$ to be determined from the joint distribution of $z_1, \dots, z_t$ as described above.

If $N(r)$ is the number of the $\binom{n}{t}$ scatter ratios in $\{R_{\xi_1 \cdots \xi_t}\}$ which are less than r we have, as in the one- and two-outlier cases,

$$\mathcal{E}(N(r)) = \binom{n}{t} P(z_1 \dots z_t < r). \qquad \dots \quad (4.5)$$

If r_α is chosen so that $\binom{n}{t} P(z_1 \dots z_t < r_\alpha) = \alpha$ then

$$P(r_t < r_\alpha) \leqslant \binom{n}{t} P(z_1 \dots z_t < r_\alpha) = \mathscr{E}(N(r_\alpha)) = \alpha. \qquad \dots \quad (4.6)$$

As a matter of fact, the probability $P(z_1 \dots z_t < r)$ can be reduced to a probability involving fewer than t independent beta variables if $t > 1$. More precisely, if t is even $P(z_1 \dots z_t < r)$ reduces to an expression involving $\frac{1}{2}t$ independent beta variables, and if t is odd it reduces to one involving $\frac{1}{2}t+\frac{1}{2}$ independent beta variables. In the case of two outliers $P(z_1 z_2 < r)$ reduces to $P(u^2 < r)$ as given by (3.3). We shall now consider the cases of three and four outliers.

In the three-outlier problem, by making use of the relation

$$\sqrt{\pi}\,\Gamma(2m) = 2^{2m-1}\Gamma(m)\,\Gamma(m+\tfrac{1}{2}) \qquad \dots \quad (4.7)$$

in (4.2) for $t = 3$, we find

$$\mathscr{E}(R^h_{\xi_1\xi_2\xi_3}) = \frac{\Gamma(n-2)\Gamma\left(\frac{n-3}{2}\right)\Gamma(n-k-2+2h)\,\Gamma\left(\frac{n-k-3}{2}+h\right)}{\Gamma(n-2+2h)\Gamma\left(\frac{n-3}{2}+h\right)\Gamma(n-k-2)\Gamma\left(\frac{n-k-3}{2}\right)}, \qquad \dots \quad (4.8)$$

from which it is seen that the distribution of $R_{\xi_1\xi_2\xi_3}$ is identical with that of the product u^2v where u and v are independent random variables having beta distributions $B_e(n-k-2, k)$ and $B_e\left(\frac{n-k-3}{2}, \frac{k}{2}\right)$, respectively. Therefore,

$$P(R_{\xi_1\xi_2\xi_3} < r) = P(u^2v < r), \qquad \dots \quad (4.9)$$

and denoting $\min\limits_{\xi_3>\xi_2>\xi_1} \{R_{\xi_1\xi_2\xi_3}\}$ by r_3

we have

$$P(r_3 < r) \leqslant \binom{n}{3} P(u^2v < r) \qquad \dots \quad (4.10)$$

where, omitting details, we find

$$P(u^2v < r) = \frac{\Gamma(n-2)\,\Gamma\left(\frac{n-3}{2}\right)}{\Gamma(n-k-2)\Gamma\left(\frac{n-k-3}{2}\right)\Gamma(k)\,\Gamma\left(\frac{k}{2}\right)} \int_0^r s^{\frac{n-k-5}{2}} \int_{\sqrt{s}}^1 (1-u)^{k-1}\left(1-\frac{s}{u^2}\right)^{\frac{k}{2}-1} du\, ds$$

$$\dots \quad (4.11)$$

For $k = 1$ this expression reduces to

$$P(u^2v < r) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-4}{2}\right)} \int_0^r s^{\frac{n-5}{2}} \int_s^1 v^{-\frac{3}{2}}(1-v)^{-\frac{1}{2}} dv\, ds \qquad \dots \quad (4.12)$$

and for $k = 2$ it reduces to

$$P(u^2v < r) = \frac{(n-3)(n-4)(n-5)}{2}(\sqrt{r})^{n-5}\left[\frac{1}{n-5} - \frac{2\sqrt{r}}{n-4} + \frac{r}{n-3}\right]. \qquad \dots \quad (4.13)$$

If we choose r_α so that $$\binom{n}{3} P(u^2v < r_\alpha) = \alpha. \quad \ldots \quad (4.14)$$

where $P(u^2v < r)$ is given by (4.10), we have

$$P(r_3 < r_\alpha) \leqslant \alpha.$$

If $N(r)$ is the number of the scatter ratios in the set $\{R_{\xi_1\xi_2\xi_3}\}$ having values $< r$ we note that $\mathcal{E}(N(r_\alpha)) = \alpha$.

In the four-outlier case by making use of (4.7) in (4.2) for $t = 4$ we find

$$\mathcal{E}(R^h_{\xi_1\xi_2\xi_3\xi_4}) = \frac{\Gamma(n-2)\Gamma(n-4)\Gamma(n-k-2+2h)\Gamma(n-k-4+2h)}{\Gamma(n-k-2)\Gamma(n-k-4)\Gamma(n-2+2h)\Gamma(n-4+2h)} \quad \ldots \quad (4.15)$$

from which we note that the distribution of $R_{\xi_1\xi_2\xi_3\xi_4}$ is identical with that of the product u^2w^2 where u and w are independent random variables having the beta distributions $B_e(n-k-2, k)$ and $B_e\ (n-k-4, k)$, respectively. Therefore

$$P(R_{\xi_1\xi_2\xi_3\xi_4} < r) = P(u^2w^2 < r) \quad \ldots \quad (4.16)$$

and denoting $\min\limits_{\xi_4>\xi_3>\xi_2>\xi_1} \{R_{\xi_1\xi_2\xi_3\xi_4}\}$ by r_4

we have $$P(r_4 < r) \leqslant \binom{n}{4} P(u^2w^2 < r) \quad \ldots \quad (4.17)$$

where, omitting details, we find that

$$P(u^2w^2 < r) = \frac{\Gamma(n-2)\Gamma(n-4)}{2\Gamma(n-k-2))\Gamma(n-k-4)\Gamma^2(k)} \int_0^r s^{\frac{n-k-6}{2}} \int_{\sqrt{s}}^1 u\left(1+\sqrt{s}-u-\frac{\sqrt{s}}{u}\right)^{k-1} du\, ds$$

$$\ldots \quad (4.18)$$

For $k = 1$ (4.18) reduces to

$$P(u^2w^2 < r) = \tfrac{1}{2}(\sqrt{r})^{n-5}[(n-3)-(n-5)r], \quad \ldots \quad (4.19)$$

and for $k = 2$ we find

$$P(u^2w^2 < r) = \frac{(n-3)!}{6(n-7)!}(\sqrt{r})^{n-6}\left[\frac{1}{n-6}-\frac{3\sqrt{r}}{n-5}+\frac{3r}{n-4}-\frac{\sqrt{r^3}}{n-3}\right]. \quad \ldots \quad (4.20)$$

Again note that if we choose k_α so that

$$\binom{n}{4} P(u^2w^2 < r_\alpha) = \alpha, \quad \ldots \quad (4.21)$$

where $P(u^2w^2 < r)$ is given by (4.18) we obtain

$$P(r_4 < r_\alpha) \leqslant \alpha. \quad \ldots \quad (4.22)$$

as in the case for $k = 3$ if $N(r)$ is the number of scatter ratios in the set $\{R_{\xi_1\xi_2\xi_3\xi_4}\}$ which have values less than r we have $\mathcal{E}(N(r_\alpha)) = \alpha$.

TABLE 2. VALUES OF r_α FOR WHICH THE UPPER BOUND $nP(u<r_\alpha)$ OF $P(r_1<r_\alpha)$ HAS THE VALUE α [OR EQUIVALENTLY. VALUES OF r_α FOR WHICH $E(N(r_\alpha))=\alpha$] FOR THE CASE OF ONE OUTLIER

$\alpha = 0.010$

sample size n	number of dimensions k: 1	2	3	4	5
5	0.02795	0.00200	0.00000		
6	.06592	.01406	.00111	0.00000	
7	.11026	.03780	.00893	.00071	0.00000
8	.15547	.06898	.02593	.00632	.00050
9	.19888	.10358	.04987	.01937	.00476
10	.23942	.13895	.07781	.03866	.01523
11	.27678	.17364	.10755	.06200	.03129
12	.31103	.20689	.13765	.08757	.05126
13	.34238	.23835	.16726	.11407	.07362
14	.37107	.26790	.19590	.14065	.09723
15	.39738	.29556	.22330	.16678	.12128
16	.42156	.32141	.24936	.19215	.14525
17	.44383	.34555	.37404	.21657	.16878
18	.46440	.36810	.29737	.23996	.19167
19	.48344	.38919	.31940	.26228	.21378
20	.50112	.40893	.34019	.28354	.23506
21	.51757	.42743	.35982	.30376	.25547
22	.53292	.44480	.37835	.32298	.27501
23	.54727	.46113	.39588	.34125	.29370
24	.56071	.47651	.41246	.35861	.31155
25	.57334	.49102	.42815	.37513	.32861
26	.58521	.50471	.44304	.39084	.34491
27	.59641	.51767	.45716	.40580	.36048
28	.60698	.52994	.47057	.42006	.37536
29	.61697	.54158	.48333	.43365	.38959
30	.62644	.55263	.49547	.44663	.40320
35	.66716	.60048	.54835	.50344	.46318
40	.69944	.63870	.59091	.54949	.51217
45	.72567	.66994	.62588	.58754	.55284
50	.74745	.69598	.65514	.61947	.58711
55	.76583	.71803	.67997	.64666	.61636
60	.78157	.73694	.70133	.67009	.64162
65	.79521	.75336	.71990	.69050	.66366
70	.80715	.76775	.73620	.70843	.68306
75	.81769	.78048	.75062	.72432	.70026
80	.82708	.79181	.76348	.73851	.71563
85	.83549	.80197	.77503	.75124	.72944
90	.84308	.81115	.78545	.76274	.74192
95	.84995	.81946	.79490	.77319	.75326
100	.85622	.82704	.80352	.78271	.76361
200	.92016	.90435	.89155	.88018	.86361
300	.94392	.93293	.92411	.91625	.90899
400	.95652	.94801	.94125	.93525	.92969
500	.96439	.95739	.95190	.94704	.94254

TABLE 2. VALUES OF r_α FOR WHICH THE UPPER BOUND $nP(u<r_\alpha)$ OF $P(r_1<r_\alpha)$ HAS THE VALUE α [OR EQUIVALENTLY, VALUES OF r_α FOR WHICH $E(N(r_\alpha)) = \alpha$] FOR THE CASE OF ONE OUTLIER—(*Continued*)

	$\alpha = 0.025$				
sample size n	number of dimensions k: 1	2	3	4	5
5	0.05124	0.00500	0.00002		
6	.10353	.02589	.00278	0.00001	
7	.15787	.05976	.01647	.00179	0.00000
8	.20934	.09953	.04111	.01166	.00125
9	.25636	.14057	.07219	.03075	.00879
10	.29873	.18053	.10601	.05606	.02420
11	.33677	.21834	.14030	.08466	.04541
12	.37094	.25361	.17380	.11452	.07008
13	.40170	.28629	.20589	.14441	.09644
14	.42950	.31647	.23627	.17360	.12331
15	.45471	.34433	.26485	.20171	.14998
16	.47768	.37007	.29165	.22854	.17601
17	.49867	.39387	.31674	.25400	.20114
18	.51794	.41593	.34022	.27811	.22525
19	.53569	.43642	.36221	.30089	.24827
20	.55208	.45547	.38281	.32239	.27020
21	.56727	.47324	.40213	.34269	.29106
22	.58139	.48984	.42028	.36187	.31088
23	.59455	.50538	.43735	.37999	.32971
24	.60685	.51996	.45343	.39713	.34760
25	.61836	.53367	.46860	.41336	.36460
26	.62917	.54657	.48292	.42873	.38076
27	.63934	.55874	.49647	.44332	.39614
28	.64891	.57025	.50930	.45717	.41079
29	.65796	.58113	.52147	.47034	.42475
30	.66651	.59144	.53303	.48287	.43806
35	.70317	.63587	.58306	.53737	.49626
40	.73208	.67113	.62301	.58117	.54333
45	.75551	.69982	.65567	.61711	.58213
50	.77492	.72365	.68286	.64715	.61466
55	.79128	.74378	.70589	.67264	.64232
60	.80527	.76102	.72564	.69454	.66613
65	.81738	.77596	.74278	.71357	.68685
70	.82798	.78904	.75780	.73027	.70504
75	.83733	.80060	.77108	.74503	.72116
80	.84566	.81088	.78291	.75820	.73553
85	.85312	.82010	.79352	.77001	.74844
90	.85984	.82841	.80308	.78067	.76009
95	.86594	.83595	.81176	.79034	.77066
100	.87150	.84281	.81967	.79916	.78030
200	.92829	.91280	.90028	.88914	.87885
300	.94949	.93871	.93008	.92240	.91530
400	.96077	.95240	.94579	.93993	.93450
500	.96783	.96093	.95555	.95082	.94642

TABLE 2. VALUES OF r_α FOR WHICH THE UPPER BOUND $nP(u<r_\alpha)$ OF $P(r_1<r_\alpha)$ HAS THE VALUE α [OR EQUIVALENTLY, VALUES OF r_α FOR WHICH $E(N(r_\alpha))=\alpha$] FOR THE CASE OF ONE OUTLIER—(*Continued*)

$\alpha = 0.050$

sample size n	number of dimensions k: 1	2	3	4	5
5	0.08083	0.01000			
6	.14529	.04110	0.00556		
7	.20661	.08452	.02620	0.00358	
8	.26161	.13133	.05831	.01856	0.00251
9	.31006	.17711	.09559	.04367	.01400
10	.35261	.22007	.13408	.07438	.03440
11	.39008	.25965	.17171	.10731	.06033
12	.42325	.29584	.20751	.14050	.08896
13	.45277	.32886	.24112	.17285	.11850
14	.47921	.35897	.27245	.20383	.14785
15	.50302	.38650	.30154	.23319	.17642
16	.52457	.41171	.32855	.26086	.20386
17	.54417	.43487	.35361	.28686	.23002
18	.56208	.45620	.37690	.31124	.25486
19	.57852	.47591	.39857	.33412	.27837
20	.59365	.49417	.41876	.35558	.30060
21	.60764	.51113	.43761	.37573	.32160
22	.62061	.52692	.45525	.39467	.34145
23	.63267	.54166	.47178	.41249	.36021
24	.64391	.55545	.48729	.42929	.37796
25	.65443	.56838	.50188	.44513	.39477
26	.66429	.58053	.51563	.46010	.41069
27	.67355	.59197	.52860	.47426	.42580
28	.68226	.60276	.54086	.48767	.44014
29	.69048	.61296	.55247	.50040	.45377
30	.69825	.62260	.56347	.51248	.46674
35	.73146	.66402	.61090	.56478	.52314
40	.75758	.69675	.64857	.60654	.56843
45	.77872	.72330	.67924	.64067	.60557
50	.79621	.74532	.70472	.66909	.63659
55	.81094	.76388	.72624	.69314	.66289
60	.82354	.77975	.74467	.71377	.68548
65	.83444	.79351	.76065	.73167	.70511
70	.84398	.80554	.77464	.74735	.72232
75	.85240	.81616	.78700	.76122	.73754
80	.85989	.82561	.79800	.77356	.75111
85	.86661	.83408	.80786	.78463	.76328
90	.87267	.84172	.81674	.79462	.77426
95	.87816	.84864	.82480	.80367	.78421
100	.88317	.85494	.83214	.81192	.79329
200	.93447	.91924	.90696	.89602	.88591
300	.95372	.94310	.93463	.92711	.92013
400	.96399	.95573	.94924	.94351	.93817
500	.97043	.96361	.95833	.95370	.94938

TABLE 2. VALUES OF r_α FOR WHICH THE UPPER BOUND $nP(u<_\alpha)$ OF $P(r_1<r_\alpha)$ HAS THE VALUE α [OR EQUIVALENTLY, VALUES OF r_α FOR WHICH $E(N(r_\alpha)) = \alpha$] FOR THE CASE OF ONE OUTLIER—(*Continued*)

$\alpha = 0.100$

sample size n	number of dimensions k: 1	2	3	4	5
5	0.10000	0.02000	0.00025		
6	.20000	.06525	.01114	0.00012	
7	.26960	.11952	.04172	.00717	0.00007
8	.32610	.17328	.08282	.02959	.00502
9	.37418	.22314	.12675	.06216	.02234
10	.41540	.26827	.16978	.09888	.04901
11	.45106	.30878	.21038	.13629	.08032
12	.48221	.34511	.24801	.17267	.11319
13	.50966	.37776	.28264	.20723	.14593
14	.53405	.40719	.31442	.23967	.17764
15	.55586	.43383	.34358	.26995	.20789
16	.57550	.45804	.37037	.29813	.23651
17	.59328	.48014	.39502	.32433	.26346
18	.60948	.50038	.41777	.34870	.28878
19	.62428	.51899	.43881	.37139	.31254
20	.63789	.53615	.45832	.39255	.33484
21	.65043	.55205	.47645	.41231	.35578
22	.66205	.56680	.49334	.43079	.37545
23	.67282	.58053	.50912	.44812	.39396
24	.68286	.59335	.52389	.46438	.41140
25	.69223	.60535	.53774	.47967	.42784
26	.70101	.61660	.55075	.49408	.44337
27	.70925	.62717	.56301	.50767	.45806
28	.71699	.63713	.57457	.52052	.47197
29	.72429	.64653	.58549	.53268	.48516
30	.73119	.65540	.59583	.54420	.49768
35	.76063	.69342	.64023	.59385	.55182
40	.78375	.72335	.67531	.63326	.59499
45	.80245	.74758	.70379	.66532	.63023
50	.81792	.76763	.72738	.69195	.65955
55	.83094	.78451	.74728	.71443	.68435
60	.84208	.79895	.76430	.73369	.70561
65	.85173	.81145	.77904	.75037	.72404
70	.86017	.82238	.79193	.76497	.74019
75	.86763	.83203	.80331	.77787	.75446
80	.87427	.84061	.81344	.78935	.76717
85	.88023	.84830	.82251	.79963	.77856
90	.88560	.85524	.83069	.80891	.78883
95	.89048	.86152	.83811	.81731	.79814
100	.89492	.86725	.84486	.82497	.80663
200	.94067	.92574	.91371	.90300	.89308
300	.95796	.94751	.93923	.93186	.92503
400	.96722	.95908	.95272	.94711	.94189
500	.97304	.96631	.96113	.95661	.95238

TABLE 3. VALUES OF $\sqrt{r_\alpha}$ FOR WHICH THE UPPER BOUND $\binom{n}{2}$ $P(u^2<r_\alpha)$ OF $P(r_2<r_\alpha)$ HAS THE VALUE α [OR EQUIVALENTLY, VALUES OF $\sqrt{r_\alpha}$ FOR WHICH $E(N(r_\alpha))=\alpha$] FOR THE CASE OF TWO OUTLIERS

$\alpha = 0.010$

sample size n	number of dimensions k: 1	2	3	4	5
5	0.03162	0.00050			
6	.08736	.01498	0.00022		
7	.14772	.04982	.00896	0.00012	
8	.20444	.09374	.03349	.00601	0.00007
9	.25544	.13926	.06744	.02449	.00433
10	.30069	.18308	.10490	.05181	.01887
11	.34076	.22397	.14265	.08337	.04152
12	.37636	.26160	.17912	.11626	.06861
13	.40812	.29606	.21363	.14889	.09761
14	.43660	.32758	.24595	.18044	.12699
15	.46228	.35641	.27605	.21050	.15589
16	.48553	.38284	.30403	.23893	.18382
17	.50670	.40713	.33002	.26568	.21057
18	.52604	.42952	.35419	.29082	.23601
19	.54380	.45019	.37667	.31441	.26013
20	.56016	.46935	.39764	.33655	.28296
21	.57528	.48714	.41721	.35735	.30454
22	.58930	.50371	.43551	.37689	.32494
23	.60234	.51918	.45266	.39528	.34423
24	.61451	.53365	.46876	.41261	.36248
25	.62588	.54722	.48390	.42895	.37975
26	.63655	.55996	.49816	.44439	.39611
27	.64657	.57196	.51162	.45899	.41164
28	.65599	.58328	.52434	.47282	.42637
29	.66489	.59397	.53637	.48594	.44037
30	.67329	.60409	.54778	.49839	.45370
35	.70925	.64753	.59694	.55228	.51159
40	.73753	.68184	.63596	.59527	.55804
45	.76043	.70968	.66772	.63038	.59611
50	.77937	.73276	.69411	.65962	.62789
55	.79532	.75222	.71638	.68436	.65482
60	.80897	.76887	.73547	.70556	.67796
65	.82077	.78328	.75201	.72396	.69804
70	.83111	.79590	.76649	.74009	.71566
75	.84023	.80704	.77928	.75434	.73124
80	.84835	.81695	.79066	.76703	.74512
85	.85563	.82583	.80086	.77840	.75757
90	.86219	.83384	.81007	.78866	.76880
95	.86814	.84110	.81841	.79797	.77899
100	.87356	.84771	.82601	.80645	.78828
200	.92902	.91518	.90349	.89291	.88304
300	94974	.94023	.93218	.92489	.91808
400	.96076	.95349	.94734	.94176	.93655
500	.96766	.96180	.95679	.95226	.94804

TABLE 3. VALUES OF $\sqrt{r_\alpha}$ FOR WHICH THE UPPER BOUND $\binom{n}{2} P(u^2 < r_\alpha)$ OF $P(r_2 < r_\alpha)$ HAS THE VALUE α [OR EQUIVALENTLY, VALUES OF $\sqrt{r_\alpha}$ FOR WHICH $E(N(r_\alpha)) = \alpha$] FOR THE CASE OF TWO OUTLIERS—(*Continued*)

$\alpha = 0.025$

sample size n	number of dimensions k: 1	2	3	4	5
5	0.05000	0.00125			
6	.11856	.02376	0.00056		
7	.18575	.06794	.01422	0.00030	
8	.24556	.11852	.04575	.00954	0.00018
9	.29758	.16820	.08546	.03346	.00687
10	.34274	.21444	.12703	.06572	.02579
11	.38212	.25661	.16755	.10110	.05269
12	.41670	.29479	.20581	.13678	.08327
13	.44728	.32932	.24141	.17138	.11495
14	.47453	.36058	.27432	.20427	.14634
15	.49896	.38897	.30468	.23522	.17670
16	.52099	.41483	.33266	.26419	.20568
17	.54097	.43848	.35849	.29124	.23314
18	.55918	.46017	.38237	.31647	.25905
19	.57585	.48014	.40450	.34003	.28346
20	.59118	.49859	.42504	.36203	.30642
21	.60532	.51567	.44415	.38260	.32802
22	.61842	.53155	.46197	.40187	.34835
23	.63058	.54633	.47863	.41995	.36751
24	.64191	.56014	.49423	.43694	.38557
25	.65250	.57307	.50887	.45292	.40262
26	.66242	.58520	.52263	.46799	.41874
27	.67173	.59660	.53560	.48221	.43399
28	.68049	.60734	.54784	.49566	.44844
29	.68874	.61748	.55941	.50839	.46215
30	.69653	.62707	.57036	.52046	.47517
35	.72985	.66814	.61742	.57252	.53152
40	.75603	.70050	.65465	.61389	.57651
45	.77720	.72671	.68487	.64757	.61326
50	.79471	.74841	.70994	.67555	.64386
55	.80946	.76669	.73108	.69919	.66975
60	.82207	.78233	.74917	.71944	.69195
65	.83300	.79586	.76484	.73699	.71121
70	.84255	.80770	.77855	.75236	.72808
75	.85099	.81815	.79066	.76593	.74299
80	.85851	.82746	.80144	.77801	.75628
85	.86524	.83579	.81109	.78884	.76818
90	.87132	.84330	.81980	.79861	.77892
95	.87683	.85012	.82769	.80746	.78866
100	.88185	.85632	.83487	.81552	.79753
200	.93335	.91971	.90818	.89774	.88801
300	.95267	.94330	.93537	.92818	.92146
400	.96298	.95582	.94976	.94426	.93912
500	.96944	.96350	.95873	.95427	.95010

TABLE 3. VALUES OF $\sqrt{r_\alpha}$ FOR WHICH THE UPPER BOUND $\binom{n}{2}$ $P(u^2<r_\alpha)$ OF $P(r_2<r_\alpha)$ HAS THE VALUE α [OR EQUIVALENTLY, VALUES OF $\sqrt{r_\alpha}$ FOR WHICH $E(N(r_\alpha))=\alpha$] FOR THE CASE OF TWO OUTLIERS—(*Continued*)

$\alpha = 0.050$

sample size n	number of dimensions k				
	1	2	3	4	5
5	0.07071	0.00250	0.00000		
6	.14938	.03372	.00111		
7	.22090	.08601	.02019	0.00060	
8	.28207	.14167	.05800	.01355	0.00036
9	.33403	.19419	.10237	.04245	.00975
10	.37841	.24188	.14702	.07881	.03273
11	.41670	.28462	.18945	.11716	.06322
12	.45006	.32286	.22884	.15489	.09657
13	.47939	.35712	.26504	.19086	.13029
14	.50539	.38792	.29819	.22462	.16313
15	.52863	.41573	.32853	.25609	.19451
16	.54952	.44096	.35634	.28532	.22417
17	.56842	.46394	.38188	.31244	.25207
18	.58562	.48496	.40540	.33763	.27823
19	.60134	.50426	.42712	.36103	.30274
20	.61578	.52205	.44722	.38282	.32571
21	.62908	.53849	.46588	.40313	.34723
22	.64139	.55374	.48324	.42210	.36743
23	.65282	.56793	.49944	.43986	.38641
24	.66346	.58117	.51459	.45651	.40426
25	.67339	.59354	.52878	.47215	.42108
26	.68269	.60515	.54211	.48687	.43695
27	.69141	.61604	.55465	.50076	.45194
28	.69962	.62630	.56648	.51386	.46612
29	.70735	.63598	.57765	.52625	.47955
30	.71465	.64513	.58821	.53799	.49230
35	.74583	.68424	.63351	.58850	.54730
40	.77032	.71501	.66926	.62850	.59106
45	.79013	.73992	.69824	.66101	.62671
50	.80652	.76052	.72224	.68798	.65635
55	.82032	.77787	.74247	.71073	.68138
60	.83213	.79271	.75978	.73021	.70284
65	.84236	.80555	.77476	.74708	.72143
70	.85132	.81678	.78787	.76185	.73772
75	.85922	.82670	.79944	.77489	.75210
80	.86627	.83553	.80974	.78650	.76491
85	.87259	.84343	.81896	.79689	.77639
90	.87829	.85056	.82728	.80627	.78674
95	.88346	.85703	.83482	.81477	.79612
100	.88817	.86292	.84169	.82251	.80467
200	.93664	.92316	.91177	.90145	.89181
300	.95490	.94564	.93780	.93069	.92405
400	.96466	.95759	.95160	.94616	.94107
500	.97080	.96500	.96021	.95580	.95167

TABLE 3. VALUES OF $\sqrt{r_\alpha}$ FOR WHICH THE UPPER BOUND $\binom{n}{2}$ $P(u^2 < r_\alpha)$ OF $P(r_2 < r_\alpha)$ HAS THE VALUE α [OR EQUIVALENTLY, VALUES OF $\sqrt{r_\alpha}$ FOR WHICH $E(N(r_\alpha)) = \alpha$] FOR THE CASE OF TWO OUTLIERS—*(Continued)*

$\alpha = 0.100$

sample size n	number of dimensions k: 1	2	3	4	5
5	0.10000	0.00501			
6	.18821	.04791	0.00223		
7	.26269	.10904	.02872	0.00119	
8	.32402	.16955	.07368	.01927	0.00072
9	.37493	.22444	.12283	.05397	.01386
10	.41780	.27305	.17039	.09468	.04161
11	.45442	.31592	.21449	.13599	.07599
12	.48609	.35381	.25473	.17565	.11219
13	.51380	.38747	.29126	.21282	.14791
14	.53826	.41753	.32440	.24728	.18212
15	.56006	.44453	.35452	.27909	.21439
16	.57962	.46892	.38196	.30842	.24461
17	.59728	.49106	.40705	.33548	.27282
18	.61332	.51125	.43006	.36047	.29911
19	.62797	.52974	.45124	.38360	.32362
20	.64140	.54676	.47079	.40506	.34649
21	.65378	.56246	.48889	.42501	.36785
22	.66522	.57700	.50571	.44360	.38783
23	.67584	.59051	.52137	.46095	.40655
24	.68572	.60310	.53598	.47720	.42412
25	.69494	.61487	.54966	.49243	.44064
26	.70357	.62589	.56250	.50675	.45619
27	.71167	.63623	.57456	.52023	.47086
28	.71928	.64596	.58592	.53293	.48472
29	.72646	.65513	.59664	.54494	.49784
30	.73323	.66380	.60677	.55631	.51026
35	.76216	.70082	.65015	.60508	.56375
40	.78489	.72990	.68431	.64361	.60614
45	.80328	.75343	.71196	.67486	.64061
50	.81850	.77288	.73485	.70074	.66921
55	.83133	.78926	.75413	.72257	.69335
60	.84231	.80327	.77061	.74124	.71401
65	.85183	.81539	.78488	.75740	.73191
70	.86017	.82600	.79736	.77155	.74758
75	.86754	.83537	.80837	.78403	.76141
80	.87410	.84370	.81817	.79514	.77372
85	.87999	.85118	.82696	.80509	.78475
90	.88531	.85791	.83488	.81407	.79470
95	.89014	.86402	.84205	.82220	.80372
100	.89454	.86960	.84859	.82961	.81193
200	.93994	.92664	.91538	.90518	.89565
300	.95713	.94799	.94025	.93322	.92666
400	.96635	.95936	.95345	.94807	.94305
500	.97215	.96650	.96169	.95733	.95326

5. Acknowledgement

The author is grateful to Mr. Paul Raynault for programming and carrying out the computations involved in Tables 2 and 3 on the IBM 7090 electronic computer. The author is also glad to acknowledge several interesting and useful discussions about multidimensional outliers with Professor Henry F. Kaiser of the University of Illinois who was also a Fellow of the Center for Advanced Study in the Behavioral Sciences in the fall of 1961.

References

David, H. A. (1956): Revised upper percentage points of the extreme studentized deviate from the sample mean. *Biometrika*, **43**, 449–452.

Dixon, W. J. (1950): Analysis of extreme values. *Ann. Math. Stat.*, **21**, 488–506.

——— (1951): Ratios involving extreme values. *Ann. Math. Stat.*, **22**, 68–78.

Ferguson, T. S. (1961): On the rejection of outliers. *Proceedings Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, University of California Press.

Grubbs, F. E. (1950): Sample criteria for testing outlying observations. *Ann. Math. Stat.*, **21**, 27–58.

Irwin, J. O. (1925): On a criterion for the rejection of outlying observations. *Biometrika*, **17**, 238–250.

Karlin, S. and Truax, D. (1960): Slippage problems. *Ann. Math. Stat.*, **31**, 296–324.

McKay, A. T. (1935): The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe. *Biometrika*, **27**, 466–471.

Nair, K. R. (1952): Tables of percentage points of studentized extreme deviate from the sample mean. *Biometrika*, **39**, 189–193.

——— (1948): The distribution of the extreme deviate from the sample mean and its studentized form. *Biometrika*, **35**, 118–144.

Newman, D. (1940): The distribution of ranges in samples from a normal population, expressed in terms of an independent estimate of the standard deviation. *Biometrika*, **31**, 20–30.

Pearson, E. S. and Chandra Sekar, C. (1936): The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika*, **28**, 308–320.

Pearson, E. S. and Hartley, H. O. (1942): Tables of probability integral of studentized ranges. *Biometrika*, **33**, 89–99.

——— (1942): The probability integral of the range in samples of n observations from a normal population. *Biometrika*, **32**, 301–310.

Pillai, K. C. S. and Tienzo, B. P. (1959): On the distribution of the extreme studentized deviate from the sample mean. *Biometrika*, **46**, 467–472.

Pillai, K. C. S. (1959): Upper percentage points of the extreme studentized deviate from the sample mean. *Biometrika*, **46**, 473–474.

Rider, P. R. (1932): Criteria for rejection of observations. *Washington University Studies*, No. 8.

Thompson, W. R. (1935): On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviates, *Ann. Math. Stat.*, **6**, 214–219.

Wilks, S. S. (1962): *Mathematical Statistics*, John Wiley and Sons, Inc., New York.

Paper received : February, 1963.

CORRIGENDA

Introducing Volume Twentyfive : By P. C. Mahalanobis, *Sankhyā*, Series A, **25,** 3. Footnote 4 should be replaced by the following :

Annual Review by P. C. Mahalanobis, Indian Statistical Institute Annual Report : 1960-61, 78-81.

Address by R. A. Fisher at the First Convocation of the Indian Statistical Institute, 12 February, 1961.

---

(1) **Approximate Probability Values for Observed Number of Successes :** By John E. Walsh, *Sankhyā*, **15**, 281-290.

(2) **Definition and Use of Generalized Percentage Points :** By John E. Walsh, *Sankhyā*, **21**, 281-288.

A quantity $C_v(x_1, x_2)$, or $C_v(n_1, n_2)$, is defined and used in the papers (1) and (2), also in *Handbook of Nonparametric Statistics : Investigation of Randomness, Moments, Percentiles, and Distributions*, D. Van Nostrand Co., 189-190, as $C_v(s_1, s_2)$. The expression for $C_v(x_1, x_2)$ is accurate for $v = 2$ but inaccurate for $v \geqslant 3$. A correct expression is

$$\sum_{j=0}^{v} (-1)^{j+1} \binom{v}{j} \sum_{x=x_1}^{x_2} p_{n-v}(x-j),$$

where $p_m(x) = \binom{m}{x} p^x q^{m-x}$ for $x = 0, 1, \ldots, m$ and is zero otherwise. The author of (1) and (2) is indebted to Frederic M. Lord for calling his attention to this discrepancy and for furnishing the correct expression. Also, as noted in (2), $\frac{1}{2}\sigma^2 C_2(x_1, x_2)$ should be replaced by $\frac{n}{2}\sigma^2 C_2(x_1, x_2)$ in the expansion on page 283 of (1).

ADDENDA

Statistics Proposed for Various Tests of Hypotheses and their Distributions in Particular Cases: By O. P. Bagai, *Sankhyā*, Series A, **24,** 409–418.

The results (4.3) and (4.4) can, respectively, be put in a standard form of the generalized Gauss' Hypergeometric series as follows :

$$\begin{aligned} I = {} & 8a^2[(1-2\gamma)-\log 4a]\ {}_0F_1(; 3 ; 4a) \\ & + \Big[1-4a+\frac{(4a)^2}{0!2!}\,\frac{1}{2}+\frac{(4a)^3}{1!\,3!}\Big(\frac{1}{3}+\frac{1}{2}+1\Big)+\frac{(4a)^4}{2!\,4!}\Big(\frac{1}{4}+\frac{1}{3}+\frac{1}{2}+1\ \Big) \\ & +\frac{(4a)^5}{3!5!}\Big(\frac{1}{5}+\frac{1}{4}+\frac{1}{3}+\frac{1}{2}+1+\frac{1}{2}+\frac{1}{3}\Big)+\ldots\Big] \end{aligned}$$

and

$$\begin{aligned} L(a) = {} & \Big[1+2\cdot\frac{a^2}{2!}\Big(\frac{1}{2}+1\Big)-\frac{2^2}{2}\,\frac{a^4}{4!}\Big(\frac{1}{4}+\frac{1}{3}+\frac{1}{2}+1+\frac{1}{2}\Big) \\ & +\frac{2^3}{2.4}\,\frac{a^6}{6!}\Big(\frac{1}{6}+\frac{1}{5}+\ldots+\frac{1}{2}+1+\frac{1}{2}+\frac{1}{4}\Big)-\ldots\Big] \\ & -a\sqrt{\pi}\ {}_0F_2\Big(\ ;\frac{1}{2},\ \frac{3}{2};\ -\frac{a^2}{4}\Big) \\ & -2(\gamma+\log a)\,\frac{a^2}{2!}\Big(;\frac{3}{2},\ 2\ ;-\ \frac{a^2}{4}\Big) \end{aligned}$$

where, in both, γ is an Euler constant and the symbol ${}_0F_p$ is defined as follows :

$${}_0F_p(; r_1, r_2, \ldots, r_p\ ; a) = 1+\frac{1}{r_1 r_2 \ldots r_p}\ \frac{a}{1!}+\frac{1}{r_1(r_1+1)\ r_2(r_2+1)\ldots r_p(r_p+1)}\,\frac{a^2}{2!}+\ldots$$

Similar change may be made in (4.16).

Distribution of the Determinant of the Sum of Products Matrix in the Noncentral Linear Case for Some Value of p : By O. P. Bagai, *Sankhyā*, Series A, **24**, 55–62.

The results (2.5), (2.8), (2.9), and (2.10) can, further, be put in a standard form of the generalized Gauss' Hypergeometric series which are, respectively, as follows :

$$I = \frac{1}{4}\Big[1-4a+\frac{(4a)^2}{0!\,2!}\cdot\frac{1}{2}+\frac{(4a)^3}{1!3!}\left(\frac{1}{3}+\frac{1}{2}+1\right)+\frac{(4a)^4}{2!4!}\left(\frac{1}{4}+\frac{1}{3}+\frac{1}{2}+1+\frac{1}{2}\right)$$

$$+\frac{(4a)^5}{3!5!}\left(\frac{1}{5}+\frac{1}{4}+\frac{1}{3}+\frac{1}{2}+1+\frac{1}{2}+\frac{1}{3}\right)+\dots\Big]+2a^2[(1-2\gamma)-\log 4a]\ {}_0F_1(;3;4a).$$

$$L_0(a) = \Gamma(1)\Big[1+2\cdot\frac{a^2}{2!}\left(\frac{1}{2}+1\right)-\frac{2^2}{2}\cdot\frac{a^4}{4!}\left(\frac{1}{4}+\frac{1}{3}+\frac{1}{2}+1+\frac{1}{2}\right)$$

$$+\frac{2^3}{2.4}\,\frac{a^6}{6!}\left(\frac{1}{6}+\frac{1}{5}+\dots+\frac{1}{2}+1+\frac{1}{2}+\frac{1}{4}\right)-\dots\Big]$$

$$-a\sqrt{\pi}\ {}_0F_2\left(;\frac{1}{2},\frac{3}{2};-\frac{a^2}{4}\right)-2(\gamma+\Gamma(1)\log a)\frac{a^2}{2!}\ {}_0F_2\left(;\frac{3}{2},2;-\frac{a^2}{4}\right).$$

$$L_1(a) = \Gamma(2)\Big[1+\frac{2}{2}\,\frac{a^2}{2!}+\frac{2^2}{2}\,\frac{a^4}{4!}\left(\frac{1}{4}+\frac{1}{3}+\frac{1}{2}+1-\frac{1}{2}\right)-\frac{2^3}{2}\,\frac{a^6}{6!}\,\frac{1}{2}\Big(\frac{1}{6}+\frac{1}{5}+\dots$$

$$+\frac{1}{2}+1-\frac{1}{2}+\frac{1}{2}\Big)+\frac{2^4}{2}\,\frac{a^8}{8\,!}\,\frac{1}{2.4}\left(\frac{1}{8}+\frac{1}{7}+\dots+\frac{1}{2}+1-\frac{1}{2}+\frac{1}{2}+\frac{1}{4}\right)-\dots\Big]$$

$$-a\Gamma\left(\frac{3}{2}\right)\ {}_0F_2\left(;-\frac{1}{2},\frac{3}{2};-\frac{a^2}{4}\right)-2\left[\left(\gamma-\frac{1}{2}\right)+\Gamma(2)\log a\right]$$

$$\frac{a^4}{4!}\ {}_0F_2\left(;\frac{5}{2},3;-\frac{a^2}{4}\right).$$

and $$L_2(a) = \Gamma(3)\Big[1+\frac{2a^2}{2!}\,\frac{1}{4}+\frac{2^2a^4}{4!}\,\frac{1}{4.2}+\frac{2^3a^6}{6!}\,\frac{1}{4.2}\left(\frac{1}{6}+\frac{1}{5}+\dots+\frac{1}{2}+1-\frac{1}{4}-\frac{1}{2}\right)$$

$$-\frac{2^4a^8}{8!}\,\frac{1}{4.2}\cdot\frac{1}{2}\left(\frac{1}{8}+\frac{1}{7}+\dots+\frac{1}{2}+1-\frac{1}{4}-\frac{1}{2}+\frac{1}{2}\right)$$

$$+\frac{2^5a^{10}}{10!}\,\frac{1}{4.2}\cdot\frac{1}{2.4}\left(\frac{1}{10}+\frac{1}{9}+\dots+\frac{1}{2}+1-\frac{1}{4}-\frac{1}{2}+\frac{1}{2}+\frac{1}{4}\right)$$

$$-\frac{2^6a^{12}}{12!}\,\frac{1}{4.2}\cdot\frac{1}{2.4.6}\left(\frac{1}{12}+\frac{1}{11}+\dots+\frac{1}{2}+1-\frac{1}{4}-\frac{1}{2}+\frac{1}{2}+\frac{1}{4}+\frac{1}{6}\right)+\dots\Big]$$

$$-a\Gamma\left(\frac{5}{2}\right)\ {}_0F_2\left(;-\frac{3}{2},\frac{3}{2};-\frac{a^2}{4}\right)-\left[\left(\gamma-\frac{3}{2}\right)+\Gamma(3)\log a\right]$$

$$\frac{a^6}{6!}\ {}_0F_2\left(;\frac{7}{2},4;-\frac{a^2}{4}\right).$$

where, in all these relations, γ is an Euler Constant and the symbol ${}_0F_p$ is defined as follows :

$${}_0F\,(;r_1,r_2,\dots,r_p;a) = 1+\frac{1}{r_1\,r_2\dots r_p}\,\frac{a}{1!}+\frac{1}{r_1(r_1+1)r_2(r_2+1)\dots(r_p)(r_p+1)}\,\frac{a^2}{2!}+\dots.$$

INDIAN STATISTICAL INSTITUTE

The Indian Statistical Institute is a Society registered in April 1932 under the Indian Societies Registration Act (Act No. XXI of 1860). Since 1 April 1960, it has been declared as an Institution of National Importance by the Indian Statistical Institute Act No. 57 of 1959.

Extracts from the Memorandum of Association

II. *The objects of the Institute shall be :—*

(1) *To promote the study and dissemination of knowledge of and research on Statistics and other subjects relating to planning for national development and social welfare.*

(2) *To provide for and undertake the collection of information, investigations, projects, and operational research for purposes of planning and the improvement of efficiency of management and production.*

(3) *To undertake any other activities which in the opinion of the Council may be usefully carried out by the Institute in the furtherance of national development and social welfare.*

IV. *The membership of the Institute and its constituent bodies shall be open to all persons irrespective of sex, nationality, race, creed or class.*

V. *The Institute shall not distribute any part of its income or profit as dividend, gifts, or bonus or in any other way among its members, provided nevertheless that office-bearers, workers and other employees shall not be debarred from receiving remuneration or participating in grants because of their also being members of the Institute.*

The Institute consists of Ordinary and Life Members, Institutional Members, Fellows, Honorary Members and Patrons. The supreme control including the power of making rules is vested in the members in General Meeting assembled. The Chairman and other office-bearers are elected once every two years. The management of the affairs of the Institute is in the hands of a Council elected annually by the members of the Institute.

Facilities are provided for student and sessional membership.

Membership fee, Rs. 25/- (foreign Rs. 30/-) per year, Institutional Membership fee, Rs. 100/- per year include subscription to one series of *Sankhyā : The Indian Journal of Statistics.*

Detailed information and Membership forms may be obtained on application to the Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta-35.

SANKHYĀ : THE INDIAN JOURNAL OF STATISTICS

EDITORS : P. C. MAHALANOBIS, C. R. RAO

Sankhyā is published in co-operation with the Indian Statistical Institute in two separate Series. *Series A* is for papers and notes on new methods and techniques. *Series B* is mainly for original data, records and applications of new methods. Each series of *Sankhyā* will contain about 400 pages per volume with plates and tables (to be issued in approximately four equal parts) which is to include communications from the Indian Statistical Institute. Papers intended to be treated as communications from the Institute should be addressed to the Journal Committee (203, Barrackpore Trunk Road, Calcutta-35) appointed by the Indian Statistical Institute, which consists of the following members :

All other papers for publication and books and reprints for review should be sent to the Editor, *Sankhyā*, 203, Barrackpore Trunk Road, Calcutta-35. Publishers are requested to send two copies of each book intended for review.

Every member of the Indian Statistical Institute will get either of the two series of the Journal free of cost. The subscription price per volume including packing and postage for each of the series for non-members, payable in advance, is : USA and Canada : \$ 10.00 ; Other Countries : £ 3 ; India : Rs. 30.00. The subscription price per volume for both the series taken together is : USA and Canada : \$ 18.00 ; Other Countries : £ 5-6s ; India : Rs. 50.00. Subscriptions should be sent to the Secretary, Statistical Publishing Society, 204/1, Barrackpore Trunk Road, Calcutta-35.

It is a condition of publication in *Sankhyā* that the paper shall not already have been published elsewhere and will not be reprinted without leave of the Editor. Authors are requested to send two typewritten copies of each manuscript. It is necessary that the raw data or measurements on which the paper is based, and detailed information of their sources, should be furnished in every case. It is desirable that the distribution or correlation charts from which statistical constants are calculated should accompany the paper. Tables should be numbered consecutively with Arabic figures without brackets (e.g. Table 1, Table 45 etc.) and should have short explanatory headings. Diagrams and drawings should also be numbered consecutively with Arabic figures and should be sent in a finished state suitable for direct reproduction with all lettering in pencil. References to tables and diagrams should be given within brackets in the text. The list of literature cited should be arranged alphabetically according to authors' names. In the case of journals the volume, number, year and page references, and in the case of printed books the name of the author and the publisher, the date of publication, and number of the edition or impression should always be stated clearly. Abbreviations, local names, vernacular measures, etc., should be written very legibly (preferably in block letters) and should be explained where they first occur in the text.

Contributors receive 25 copies (joint authors : 15 copies each) of their papers free. The cost of additional reprints will depend upon the number required ; for every 50 additional copies the rate is Rs. 12/- for each sheet of eight pages or Rs. 2/- per page for part of a sheet, with an extra charge for plates printed on special paper. Additional copies should be ordered at the time of returning the final proof.

Publication of papers, notes, reviews, discussion, etc. does not signify the Editor's approval of the arguments used or conclusions drawn by the authors. It is to be distinctly understood that all opinions expressed are purely individual and no responsibility for such opinions rests on the Editor, the Journal Committee or the Indian Statistical Institute.

Printed by Kalipada Mukherjee at EKA PRESS, 204/1 Barrackpore Trunk Road, Calcutta-35 and published by Anikendra Mahalanobis from Statistical Publishing Society, 204/1 B. T. Road, Calcutta-35